Mémoire présenté pour obtenir

L'HABILITATION A DIRIGER DES RECHERCHES

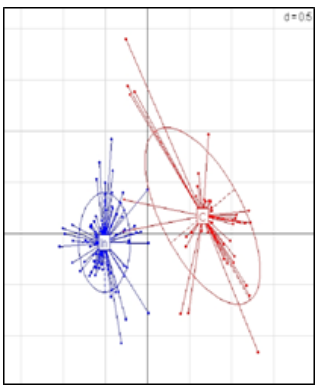
UNIVERSITE DE MONTPELLIER
Ecole Doctorale GAIA (N° 584)
(Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau)

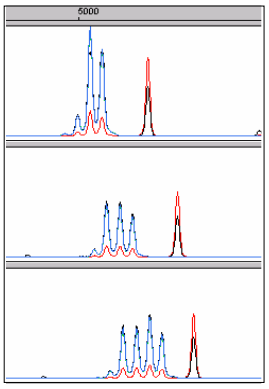
Acquisition de connaissances approfondies sur la génétique de l'espèce *Dioscorea alata* pour la production de variétés à haut rendement et de qualité

par

Gemma ARNAU

Chercheuse au Cirad
UMR AGAP Institut (UMR 1334 « Amélioration Génétique et Adaptation des Plantes méditerranéennes et tropicales »)





Soutenance **le 2 juillet 2025** - Membres du Jury :

Mme Karine ALIX, rapporteure, Professeure des universités, AgroParisTech - Université Paris-Saclay
M. Mathieu ROUSSEAU-GUEUTIN, rapporteur, Directeur de recherches, INRAE, UMR IGEPP
Mme Bernadette JULIER, rapporteure, Directrice de recherches, INRAE, URP3F
M. Jacques DAVID, examinateur, Professeur des universités, Institut Agro Montpellier, UMR AGAP
Mme Nathalie CHANTRET, examinatrice, Directrice de recherches, INRAE, UMR AGAP

Je déclare avoir respecté, dans la conception et la rédaction de ce mémoire d'HDR, les valeurs et principes d'intégrité scientifique destinés à garantir le caractère honnête et scientifiquement rigoureux de tout travail de recherche, visés à l'article L.211-2 du Code de la recherche et énoncés par la Charte nationale de déontologie des métiers de la recherche et la Charte d'intégrité scientifique de l'Université de Montpellier. Je m'engage à les promouvoir dans le cadre de mes activités futures d'encadrement de recherche.

Table des matières

1. Introduction

J’ai été recrutée par le CIRAD en 2001, en tant que responsable du programme de sélection et d'amélioration variétale de l'igname (*Dioscorea alata*) mené en Guadeloupe, fonction que j'ai occupée jusqu'en 2020. Mes activités étaient initialement centrées sur la région Caraïbe. Cependant, au fil des années, j'ai eu l'opportunité d'élargir mes partenariats à une échelle internationale, en collaborant notamment avec le Central Tuber Crops Research Institute (CTCRI) en Inde et plusieurs pays d'Afrique de l'Ouest. Cette ouverture internationale a considérablement enrichi la portée de mes travaux, renforçant ainsi les collaborations autour de problématiques communes liées à la production et à l'amélioration des ignames.

Mes recherches se sont essentiellement articulées autour de trois thématiques principales, toutes visant à accroître l'efficacité des programmes d’amélioration génétique. Ces axes incluent : (1) l'acquisition de connaissances sur la structure et le fonctionnement du génome de *Dioscorea alata*, (2) l'exploration et la valorisation de la diversité génétique disponible, et (3) l'étude du déterminisme génétique des caractères d'intérêt. Les résultats les plus marquants obtenus dans chacun de ces domaines constituent le cœur des travaux que je présente dans ce mémoire.

Mon parcours m'a également permis de collaborer avec de nombreux collègues — chercheurs, ingénieurs, techniciens et étudiants. Je tiens à exprimer ma profonde gratitude à tous ces collaborateurs, qu'ils soient issus du CIRAD ou de diverses institutions partenaires, telles que l'INRAE, le CTCRI en Inde, le MANDR en Haïti, l'IITA au Nigeria, le CNRA en Côte d'Ivoire et l’Université de Sherbrooke au Canada, entre autres.

2.Curriculum vitae

Gemma ARNAU

Chercheuse au CIRAD (depuis 2001)
Née le 09-02-1965 à Barcelone
1 enfant
Nationalité : Française

Adresse :
CIRAD, UMR AGAP
Campus Agropolis - Bâtiment ARCAD, Bureau 107
Rue Arthur Young, 34000 Montpellier Cedex 1
Téléphone : 04 32 72 29 89
Email : gemma.arnau@cirad.fr
ORCID : 0000-0003-0714-6830

DIPLOMES

1996 : Thèse de doctorat, École Nationale Agronomique de Montpellier, mention très honorable, appellation Doctorat Européen, Directeur de thèse : Prof. A. Charrier (France)

1991 : DEA de Bases de la Production Végétale : Option Biotechnologie et Amélioration des Plantes, Université Montpellier II/ENSA Montpellier (France)

1989 : Maîtrise en Sciences Biologiques, spécialisation Physiologie Végétale, Université de Barcelone (Espagne)

COMPETENCES LINGUISTIQUES

Français, Anglais, Catalan, Espagnol : courant

EXPÉRIENCE PROFESSIONNELLE

CIRAD Montpellier : 2021-présent

- Équipe DEFI (Déterminisme, Expression et Fonctionnement des traits d'intérêt chez l'Igname), dirigée par Hana Chair
- Projet de recherche : Compréhension de l'architecture génétique des traits d'intérêt et de la biologie de la reproduction chez l'igname

CIRAD Guadeloupe : 2002-2020

- Responsable du programme d'Amélioration Génétique de l'Igname
- Activités principales :
    - Direction des activités de création et de sélection variétale en collaboration étroite avec les utilisateurs finaux

- Recherche centrée sur la biologie de la reproduction des ignames et le déterminisme génétique des traits d'intérêt pour améliorer l'efficacité des programmes d'amélioration

Université de Barcelone, Espagne : 2000

- Chercheur post-doctoral, Faculté de Biologie, Université de Barcelone
- Équipe dirigée par Jose Louis Araus
- Projet de recherche : Mise en place d'un laboratoire de marquage moléculaire pour la caractérisation des ressources génétiques de Blé dur

GEVES (Groupe d'Étude et de Contrôle des Variétés et des Semences), Surgères : 1997-1999

- Chercheur post-doctoral
- Projet de recherche : Développement de techniques de marquage moléculaire (STS, AFLP, ISSR) pour l'étude de la diversité génétique et l'identification variétale de différentes espèces ornementales et potagères

STAGES

- **VARTC, Vanuatu (**Pacifique Sud, 6 mois, 2001) et **IITA, Bénin** (3 mois, 2002) : Initiation à la culture des ignames et collecte d'échantillons.

COORDINATION DE PROJETS

- **Coordination du projet franco-indien CEFIPRA 3000-B1, CIRAD-CTCR**I : " Utilisation de marqueurs d'ADN pour l'amélioration génétique de la productivité, la capacité de stockage et la teneur en matière sèche des tubercules d'igname" (2005-2010, 70,9 k€).
- **Coordination des recherches sur l'igname dans le cadre des projets FEDER** (Union Européenne - Région Guadeloupe) :
  - VALEXBIOTROP (2008-2014, 600 k€)
  - CARAMBA (2015-2017, 150 k€)
  - CAVALBIO : "Caractérisation et valorisation de la biodiversité végétale tropicale d'intérêt agronomique" (phase 2015-2017, 750 k€ ; phase 2018-2019, 325 k€)
- **Coordination des activités d'innovation variétale dans le cadre des projets FEADER** (Fonds européen agricole pour le développement rural) sur l'igname :
  - Projet DOCUP 1 (2006-2007, 330 k€)
  - Projet DOCUP 2 (2008-2011, 274 k€)
  - RITA (2012-2015, 400 k€)
  - PRODIMAD 1 (2016-2018, 265 k€)
  - PRODIMAD 2 (2019-2020, 105 k€)
- **Coordination des travaux sur l'igname dans le cadre du projet INTERREG-DEVAG** : "Un réseau régional pour le développement de systèmes de culture agroécologiques pour les cultures horticoles dans les Caraïbes" (2010-2013, 65 k€).
- **Coordination des recherches conduites en Guadeloupe dans le cadre du projet international AFRICAYAM** (8 pays, 12 partenaires) : "Amélioration de l'igname pour

une productivité accrue et une qualité améliorée en Afrique de l'Ouest" (2015-2019, 267 k€).

COLLABORATIONS FRUCTUEUSES

- **Projet franco-indien CEFIPRA (resp. K. Abraham & G. Arnau), 2005-2010** : L'objectif principal était de développer des marqueurs moléculaires pour optimiser l'amélioration génétique des ignames, en particulier en termes de productivité, qualité, capacité de stockage et teneur en matière sèche des tubercules. J'ai eu l'opportunité de collaborer étroitement avec le Dr. K. Abraham, une figure pionnière dans l'amélioration des ignames *Dioscorea alata*. Au sein de ce projet, j'ai co-encadré une doctorante, A. Nemorin, en collaboration scientifique avec J. David. Plusieurs articles ont été publiés (Nemorin *et al*., 2012 ; Abraham *et al*., 2013 ; Nemorin *et al*., 2013 ; Sheela *et al*., 2016).

- **Projet international AFRICAYAM, 2015-2019** : Ce projet avait pour but de renforcer les compétences en phénotypage, gestion des essais, gestion de programmes, et d'introduire des méthodes innovantes de sélection dans les principaux pays producteurs d'ignames en Afrique de l'Ouest. J'ai encadré un doctorant de Côte d'Ivoire, E. Ehounou, en co-encadrement scientifique avec A.M. Kouakou. Plusieurs articles ont été publiés (Ehounou *et al*., 2021 ; Ehounou *et al*., 2022 ; Arnau *et al.*, 2023).

FORMATIONS

J'ai dispensé plusieurs formations destinées à des chercheurs, doctorants ou ingénieurs.

• **2006 (21 jours) :** Formation sur la caractérisation des ressources génétiques indiennes à l'aide de marqueurs mirosatellites, par multiplexage et analyse sur séquenceur capillaire. Cette formation, destinée à la chercheuse indienne Dr. Sheela, partenaire du projet CEFIPRA, s'est tenue en Guadeloupe.

• **2006 (11 jours)** : Initiation à l'extraction d'ADN de haute qualité par purification des noyaux cellulaires, destinée aux partenaires du projet CEFIPRA en Inde.

• **2008 (20 jours)** : Formation sur la caractérisation de la diversité génétique des ignames *D. alata* cubaines et sa comparaison avec la diversité mondiale à l'aide de marqueurs SSRs. Formation dispensée au doctorant cubain M. Borges, en Guadeloupe.

• **2010 (20 jours)** : Initiation à l'amélioration génétique des ignames, destinée à l'ingénieur agronome haïtien C.D Joseph partenaire du projet INTERREG, en Guadeloupe.

• **2018 (6 mois)** : Initiation à la Spectroscopie Proche Infrarouge (SPIR) comme méthode de sélection pour la qualité du tubercule, destinée au doctorant ivoirien E. Ehounou, en Guadeloupe.

• **2019 (6 mois)** : Formation à la détection de QTLs (Quantitative Trait Loci) à partir de populations en ségrégation, destinée au doctorant ivoirien E. Ehounou, en Guadeloupe.

ENCADREMENT

Licence

1. **E. Boulerne (7 janvier—19 avril 2019, Université de Sherbrooke, Québec)** : "Phénotypage des tubercules de populations biparentales de *Dioscorea alata* L. pour l'identification de QTLs liés à la qualité."
2. **A.M. Lafontaine (7 janvier—19 avril 2019, Université de Sherbrooke)** : "Phénotypage des tubercules de populations biparentales de *Dioscorea alata* L. pour l'identification de QTLs liés à la qualité."
3. **J.F. Nadeau (septembre—décembre 2016, Université de Sherbrooke)** : "Contribution à la cartographie du génome de l'igname *Dioscorea alata* à l'aide de marqueurs SNPs et microsatellites."
4. **M.F. Bossanyi (septembre—décembre 2016, Université de Sherbrooke)** : "Contribution à la cartographie du génome de l'igname *Dioscorea alata* à l'aide de marqueurs SNPs et microsatellites."
5. **D. Blouin (janvier—avril 2012, Université de Sherbrooke)** : "Processus de cartographie génétique : lecture des résultats du logiciel Genemapper et analyses de liaisons et de cartographie génétique à l'aide du logiciel JoinMap."
6. **G. Satre (janvier—avril 2012, Université de Sherbrooke**) : "Processus de cartographie génétique : lecture des résultats du logiciel Genemapper et analyses de liaisons et de cartographie génétique à l'aide du logiciel JoinMap."
7. **A. Amiot (septembre—décembre 2011, Université de Sherbrooke)** : "Analyse de la diversité génétique des principales collections mondiales d'ignames *Dioscorea alata* à l'aide d'un set de marqueurs microsatellites commun."
8. **P.E. Lepage (mai—juillet 2011, Université de Sherbrooke)** : "Construction d'une carte génétique chez *Dioscorea alata* au niveau tétraploïde à l'aide de marqueurs microsatellites."
9. **G. Clement (mai—juillet 2011, Université de Sherbrooke)** : "Analyse de la diversité génétique des principales collections mondiales d'ignames *Dioscorea alata* à l'aide d'un set de marqueurs microsatellites commun."
10. **J. Messiva (février 2011, Université des Antilles et de la Guyane, Guadeloupe**) : "La culture in vitro d'ignames en appui aux travaux d'amélioration génétique."
11. **F. Fanton (mai—juillet 2010, Université de Sherbrooke)** : "Analyse de la diversité génétique des principales collections mondiales d'ignames *Dioscorea alata* à l'aide d'un set de marqueurs microsatellites commun."
12. **A. Paulo (mai—juillet 2009, Université de Sherbrooke)** : "Caractérisation des ressources génétiques chez l'igname *Dioscorea* sp."
13. **M. Vallée (septembre—décembre 2007, Université de Sherbrooke)** : "Mise au point d'une méthodologie de lyophilisation du pollen chez l'igname *Dioscorea* sp."
14. **I. Bacchand (mai—juillet 2006, Université de Sherbrooke)** : "Génotypage et études d'hérédité sur la collection d'ignames *Dioscorea alata* à l'aide de marqueurs moléculaires."
15. **S. Côté (mai—juillet 2006, Université de Sherbrooke**) : "Génotypage et études d'hérédité sur la collection d'ignames *Dioscorea alata* à l'aide de marqueurs moléculaires."

16. **I. Bacchand (septembre—décembre 2005, Université de Sherbrooke)** : "Génotypage de la collection d'ignames *Dioscorea alata* et recherche de marqueurs moléculaires liés à la résistance à l'anthracnose.

Master 2

1. **Maroua Zair (2024, Université de Clermont-Ferrand)** : " *Cartographie des QTLs (Quantitative Trait Loci) et recherche de gènes candidats associés à la teneur en amylose et en amylopectine chez l'igname Dioscorea alata*". Actuellement étudiante en Master en Analyses statistiques.
2. **Julia Castro (2018, Université de Barcelone)** : "*"Contribution à la création d'une carte génétique de référence de Dioscorea alata"*. Actuellement chercheuse post-doctorale à l'Université de Washington à St. Louis.

Doctorants

1. **Adou Emmanuel Ehounou (2017—2019)** : *"Détermination de QTL liés à la morphologie des tubercules et prédiction de la qualité de l'igname pilée de l'espèce Dioscorea alata"*, co-encadrement avec Amani Michel Kouakou. Actuellement chercheur au Centre National de Recherche Agronomique (CNRA) de Bouaké, Côte d'Ivoire.
2. **Alice Némorin (2009—2012)** : *"Acquisition de connaissances sur la génétique de l'espèce Dioscorea alata L. pour la production de variétés polyploïdes"*, co-encadrement avec Jacques David. Actuellement enseignante de SVT à l'Académie de Guadeloupe.
3. **Misterbino Borgès García (2009—2011)** : *"Caractérisation de la diversité génétique et diagnostic viral de Dioscorea alata L. en vue de sa micropropagation et conservation à Cuba"*, co-encadrement avec Amadou Bâ. Actuellement professeur à l'Université de Granma, Cuba.

PUBLICATIONS & COMMUNICATIONS

28 publications à comité de lecture (8 en premier auteur, 7 en dernier auteur)

3 chapitres de livre

33 communications à des conférences internationales

Publications indexées

Journaux	Nombre d'articles	Facteur d'impact (2022)
Molecular Ecolgy Resources	1	8,6
Theoretical and Applied Genetics	4	5,4
BMC Plant Biology	1	5,3
Scientific Reports	1	4,9
Annals of Botany	2	4,2

Plants, People, Planet	1	4,17
Journal of the Science of Food and Agriculture	2	4,1
PLoS One	1	3,7
Molecular Breeding	1	3,1
Photosynthetica	1	2,7
Ecology and Evolution	1	2,6
Genetic Resources and Crop Evolution	2	2
Euphytica	1	1,9
Journal of Near Infrared Spectroscopy	1	1,8
Journal of Genetics and Breeding	1	1,4

Indicateurs de citation: h-index : 13 (Scopus, le 22/11/24) ; Nombre de citations : 693

Publications

1. Houngbo, M. E., Desfontaines, L., Diman, J. L., **Arnau, G.**, Mestres, C., Davrieux, F., Rouan, L., Beurier, G., Marie-Magdeleine, C., Meghar, K., & Cornet, D. (2024). Convolutional neural network allows amylose content prediction in yam (*Dioscorea alata* L.) flour using near infrared spectroscopy. *Journal of the Science of Food and Agriculture*, 104(8), 4915-4921.
2. **Arnau, G.**, Desfontaines, L., Ehounou, A. E., Marie-Magdeleine, C., Kouakou, A. M., Leinster, J., Nudol, E., Maledon, E., & Chair, H. (2024). Quantitative trait loci and candidate genes for physico-chemical traits related to tuber quality in greater yam (*Dioscorea alata* L.). *Journal of the Science of Food and Agriculture*, 104(8), 4872-4879.
3. **Arnau, G.**, Ehounou, A. E., Maledon, E., Nudol, E., Vignes, H., Gravillon, M. C., N'Guetta, A. S. P., Mournet, P., Hâna, H., & Cormier, F. (2023). Identification and validation of QTLs for tuber quality related traits in yam *Dioscorea alata*. *Acta Horticulturae*, 1362, 381-388.
4. Kouakou, A. M., Chair, H., Dibi, K. E. B., Dossa, K., **Arnau, G.**, Ehounou, A. E., & Cornet, D. (2023). Advancing breeding for climate-resilient yam production in Côte d'Ivoire. *Plants, People, Planet*, 2023.
5. Ehounou, A. E., Cormier, F., Maledon, E., Nudol, E., Vignes, H., Gravillon, M. C., N'Guetta, A. S. P., Mournet, P., Chair, H., Kouakou, A. M., & **Arnau, G.** (2022). Identification and validation of QTLs for tuber quality related traits in greater yam (*Dioscorea alata* L.). *Scientific Reports*, 12(1), 8423.
6. Ehounou, A. E., Cornet, D., Desfontaines, L., Marie-Magdeleine, C., Maledon, E., Nudol, E., Beurier, G., Rouan, L., Brat, P., Lechaudel, M., Nous, C., N'Guetta, A. S. P., Kouakou, A. M., & **Arnau, G.** (2021). Predicting quality, texture and chemical content of yam (*Dioscorea alata* L.) tubers using near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 29(3), 128-139.
7. Cormier, F., Martin, G., Vignes, H., Lachman, L., Cornet, D., Faure, Y., Maledon, E., Mournet, P., **Arnau, G.**, & Chair, H. (2021). Genetic control of flowering in greater yam (*Dioscorea alata* L.). *BMC Plant Biology*, 21, 1-12.

8. Sharif, B. M., Burgarella, C., Cormier, F., Mournet, P., Causse, S., Nguyen Van, K., Kaoh, J., Rajaonah, M. T., Lakshan, S. R., Waki, J., Bhattacharjee, R., Badara, G., Pachakkil, B., Pavis, C., **Arnau, G.,** & Chair, H. (2020). Origin and diversification of the polyploid and clonally propagated greater yam (*Dioscorea alata* L.). *Annals of Botany*, 6(2), 1029-1038.
9. Cormier, F., Lawac, F., Maledon, E., Gravillon, M. C., Nudol, E., Mournet, P., Vignes, H., Chair, H., & **Arnau, G.** (2019). A reference high-density genetic map of greater yam (*Dioscorea alata* L.). *Theoretical and Applied Genetics*, 132(6), 1733-1744.
10. Cormier, F., Mournet, P., Causse, S., **Arnau, G.,** Maledon, E., Gomez, R. M., Pavis, C., & Chair, H. (2019). Development of a cost-effective single nucleotide polymorphism genotyping array for management of greater yam germplasm collections. *Ecology and Evolution*, 9(10), 5617-5636.
11. Ehounou, A. E., Kouakou, A. M., N'Zi, J. C., Dibi, K. E. B., Bakayoko, Y., Essis, B. S., N'Zue, B., **Arnau, G**., Maledon, E., Asfaw, A., Adebola, P., & N'Guetta, A. S. P. (2019). Production of hybrid seeds by intraspecific crossing in yam *Dioscorea alata* L. *International Journal of Science and Research (IJSR)*, 8(9), 1212–1221.
12. García, M. B., Kosky, R. G., Rodríguez, S. M., Pupo, J. J. S., Abeal, E. E., Avalos, D. R., Jerez, Y. H., Paneque, O. G., Malaurie, C. B., Hamon, C. P., Scarcelli, C. N., **Arnau, G**., Ba, C., Noyer, J. L., & Filloux, D. (2018). Caracterización de la diversidad genética de *Dioscorea alata* L. y optimización de la producción de plantas in vitro como fuente de semilla en Cuba. *Anales de la Academia de Ciencias de Cuba*, 8(1).
13. **Arnau, G**., Bhattacharjee, R., Chair, H., Malapa, R., Lebot, V., Abraham, K., Perrier, X., Petro, D., Penet, L., & Pavis, C. (2017). Understanding the genetic diversity and population structure of yam (*Dioscorea alata* L.) using microsatellite markers. *PloS One*, 12(3), e0174150.
14. Sheela, M. N., Abhilash, P. V., Asha, K. I., & **Arnau, G.** (2016). Genetic diversity analysis in greater yam (*Dioscorea alata* L.) native to India using morphological and molecular markers. *Acta Horticulturae*, 1118, 51-58.
15. Nemorin, A., David, J., Maledon, E., Nudol, E., Dalon, J., & **Arnau, G**. (2013). Microsatellite and flow cytometry analysis to understand the origin of *Dioscorea alata* polyploids. *Annals of Botany*, 112(5), 811-819.
16. Abraham, K., Nemorin, A., Lebot, V., & **Arnau, G.** (2013). Meiosis and sexual fertility of autotetraploid clones of greater yam *Dioscorea alata* L. *Genetic Resources and Crop Evolution*, 60(3), 819-823.
17. Nemorin, A., Abraham, K., David, J., & **Arnau, G.** (2012). Inheritance of tetraploid *Dioscorea alata* and evidence of double reduction using microsatellite marker segregation analysis. *Molecular Breeding*, 30(4), 1657-1667.
18. Fernandez, P., **et al**. (2011). The INTERREG-DEVAG project: a regional network for the development of agroecological cropping systems for horticultural crops in the Caribbean. *Acta Horticulturae*, 894, 147-151.
19. Bussière, F., Cabidoche, Y. M., Pétro, D., Sierra, J., Cornet, D., Guyader, S., Ozier-Lafontaine, H., Tournebize, R., **Arnau, G**., & Pavis, C. (2011). Des innovations pour les enjeux multiples des productions vivrières et maraîchères des Antilles. *Innovations Agronomiques*, 16, 39-51.
20. Andris, M., Aradottir, G. I., **Arnau, G**., et al. (2010). Permanent genetic resources added to molecular ecology resources database 1 June 2010 - 31 July 2010. *Molecular Ecology Resources*, 10(6), 1106-1108.

21. Arnau, G., Nemorin, A., Maledon, E., & Abraham, K. (2009). Revision of ploidy status of *Dioscorea alata* L. (Dioscoreaceae) by cytogenetic and microsatellite segregation analysis. *Theoretical and Applied Genetics*, 118(7), 1239-1249.
22. Bousalem, M., Arnau, G., Hochu, I., Arnolin, R., Viader, V., Santoni, S., & David, J. (2006). Microsatellite segregation analysis and cytogenetic evidence for tetrasomic inheritance in the American yam *Dioscorea trifida* and a new basic chromosome number in the Dioscoreae. *Theoretical and Applied Genetics*, 113, 439-451.
23. Malapa, R., Arnau, G., Noyer, J. L., & Lebot, V. (2005). Genetic diversity of the greater yam (*Dioscorea alata* L.) and relatedness to *D. nummularia* Lam. and *D. transversa* Br. as revealed with AFLP markers. *Genetic Resources and Crop Evolution*, 52, 919-929.
24. Baranger, A., Aubert, G., Arnau, G., Laine, A. L., Deniot, G., Potier, M., Weinachter, C., Lejeune-Henaut, A., & Lallemand, J. (2004). Genetic diversity within *Pisum sativum* using protein- and PCR-based markers. *Theoretical and Applied Genetics*, 108, 1309-1321.
25. Arnau, G., Lallemand, J., & Bourgoin, M. (2002). Fast and reliable strawberry cultivar identification using inter-simple sequence repeat (ISSR) amplification. *Euphytica*, 129, 69-79.
26. Arnau, G., Lallemand, J., & Bourgoin, M. (2001). Are AFLP markers the best alternative for cultivar identification? *Acta Horticulturae*, 546, 301-306.
27. Arnau, G., Monneveux, P., This, D., & Alegre, L. (1997). Photosynthesis of six barley genotypes as affected by water stress. *Photosynthetica*, 34, 67-76.
28. Arnau, G., & Monneveux, P. (1995). Physiology and genetics of terminal water stress tolerance in barley. *Journal of Genetics and Breeding*, 49, 327-332

Chapitres d'ouvrages

1. Chaïr H., Arnau G., Zotta Mota A. (2022). Yam genomics. In : *Underutilised crop genomes*, Chapman Mark A. (ed.). Cham : Springer, 373-389. (Compendium of Plant Genomes).
2. Arnau G., Abraham K., Sheela M.N., Hâna C., Sartie A., Robert A. (2010). Yams. In : *Root and tuber crops*, Bradshaw John E. (ed.). New York : Springer [États-Unis], 127-147. (Handbook of Plant Breeding).
3. Rekika D., Arnau G., El-Jaafari S., Monneveux P. (1995). Photosynthetic Gas Exchange Parameters as Predictive Criteria for Drought Resistance in Durum Wheat and Barley. In : *Photosynthesis: from light to biosphere*, P. Mathis (ed.), Kluver Academic Publishers, Netherlands, Vol. IV, 721-724.

Communications à des conferences Internationales

1. Arnau G., Ehounou A.E., Erick M., Nudol E., Vignes H., Gravillon M.C., N'Guetta A.S.P., Mournet P., Chaïr H., Cormier F. (2022). Identification and validation of QTLs for tuber quality related traits in yam *Dioscorea alata*. *International Horticultural Congress (IHC 2022): International Symposium on Breeding and Effective Use of Biotechnology and Molecular Tools in Horticultural Crops*, Angers, France, 14-22 Août.
2. Ehounou A.E., Cormier F., Cornet D., Maledon E., Desfontaines L., Marie-Magdeleine C., Nudol E., Gravillon M.C., Kouakou A.M., Chaïr H., Arnau G. (2019). Development of NIRS and molecular marker to improve breeding efficiency in Greater Yam (*Dioscorea*

alata L.) for key quality traits. *International Conference on Applied Biochemistry and Biotechnology (ABB 2019)*, 21-24 Juillet, Macau, Chine.

3. Cornet D., Desfontaines L., Cormier F., Marie-Magdeleine C., **Arnau G.**, Meghar K., Davrieux F., Beurier G. (2019). Assembler la diversité des modèles classiques et "deep learning" pour développer un pipeline de calibration SPIR performant et générique. *20èmes Rencontres HélioSPIR*, 14-15 Octobre, Montpellier, France.
4. Ehounou A.E., Erick M., Fabien C., Cornet D., Nudol E., Kouakou A.M., Hâna H., **Arnau G.** (2018). Breeding for improved tuber quality in greater yam (*Dioscorea alata L.*). *Triennial Symposium of the International Society for Tropical Root Crops (ISTRC)*, 22-25 Octobre, Cali, Colombie.
5. Chaïr H., Bhattacharjee R., Pavis C., Summo M., Cormier F., **Arnau G.**, Lebot V. (2018). Greater yam (*Dioscorea alata L.*) pre-breeding and breeding: use of genomic tools to decipher the genetic diversity and identify wild relatives. *Plant and Animal Genome XXVI Conference (PAG)*, 13-17 Janvier, San Diego, États-Unis.
6. **Arnau G.,** Bhattacharjee R., Sheela M.N., Chaïr H., Malapa R., Lebot V., Penet L., Pavis C. (2017). First genetic diversity analysis in greater yam (*Dioscorea alata L.*) of a representative world germplasm using microsatellite markers. *Symposium Internacional de Raices, Rizomas, Tuberculos, Platanos y Papaya*, INIVIT, 24-27 Octobre, Varadero, Cuba.
7. **Arnau G.,** Maledon E., Nudol E., Gravillon M.C. (2016). Progress and challenges in genetic improvement of yam (*Dioscorea alata L.*). *First World Congress on Root and Tuber Crops (WCRTC)*, 18-24 Janvier, Nanning, Guangxi, Chine.
8. **Arnau G.**, Lopez-Montes A., Maledon E., Nudol E. (2016). Meta-QTL analysis of the genetic control of quality related traits in Yam (*Dioscorea alata* L.) [Poster]. First World Congress on Root and Tuber Crops (WCRTC), 18-24 Janvier, Nanning, Guangxi, Chine.
9. Chaïr H., Scarcelli N., **Arnau G.**, Pavis C., Akakpo R., Dansi A., Petro D., Alix K., Lebot V., Vigouroux Y. (2016). CIRAD, IRD and INRA Yam genomic initiatives: unlocking genetic diversity and accelerating yam breeding. *Plant and Animal Genome Conference XXIV*, 9-13 Janvier, San Diego, États-Unis.
10. Chaïr H., Scarcelli N., **Arnau G.,** Pavis C., Akakpo R., Dansi A., Petro D., Alix K., Lebot V., Vigouroux Y. (2016). CIRAD, IRD and INRA Yam genomic initiatives: unlocking genetic diversity and accelerating yam breeding. *XXIV International Plant and Animal Genome Conference*, 9-13 Janvier, San Diego, Californie, États-Unis.
11. Champoiseau P., Lévy L., Osseux J., Petro D., Tournebize R., **Arnau G.**, Maledon E., Nudol E., Cornet D. (2016). A Yam collaborative selection platform in Guadeloupe: a model for effective multipartnerial and participative program. *52nd Congress of the Caribbean Food Crop Society (CFCS)*, 10-16 Juillet, Gosier, Guadeloupe.
12. **Arnau G.,** Nemorin N., Cornet D., Maledon E., Nudol E. (2013). Recent innovations in yam *Dioscorea alata* improvement through polyploidy breeding. *First World Conference on Yam*, 1-7 Octobre, Accra, Ghana.
13. **Arnau G.**, Némorin A., Maledon E., Lambert F., Nudol E. (2013). Morpho-agronomic, cytogenetic and molecular characterization of the CIRAD yam collection for their enhancement and utilization in genetic improvement programs [Poster]. *First Global Conference on Yam*, 1-7 Octobre, Accra, Ghana.
14. **Arnau G.**, Maledon E., Prophète E., Scutt R., Joseph C.D. (2013). Introduction de nouvelles variétés d'ignames en Haïti : résultats et perspectives. *Conférence finale du*

projet DEVAG Réseau caribéen pour le développement de systèmes horticoles agroécologiques, CIRAD.

15. Arnau G., Nemorin A., Maledon E., Nudol E. (2011). Advances on polyploid breeding in yam *D. alata*. *First International Symposium on Roots, Rhizomes, Tubers, Plantains, Bananas and Papaya*, 7-10 Novembre, Santa Clara, Cuba.
16. Fernandes P. et al. (2011). The INTERREG-DEVAG project: a regional network for the development of agroecological cropping systems for horticultural crops in the Caribbean. [Poster]. *Assuring Caribbean Food and Nutrition Security in the Context of Climate Change*, 47th Annual Meeting Caribbean Food Crops Society (CFCS), 3-9 Juillet, Bridgetown, Barbade.
17. Fernandes P. et al. (2010). The INTERREG-DEVAG project: a regional network for the development of agroecological cropping systems for horticultural crops in the Caribbean. *First International Symposium on Tropical Horticulture (TropHort)*, 22-26 Novembre, Kingston, Jamaïque.
18. Abraham K., Némorin A., Lebot V., Sheela M.N., Sreekumari M.T., Arnau G. (2009). Polyploidy breeding in greater yam for tuber yield improvement [Abstract]. *International Conference on Polyploidy, Hybridization and Biodiversity*, 17-20 Mai, Saint-Malo, France.
19. Arnau G., Némorin A., Maledon E., Abraham K. (2009). Revision of ploidy levels of *Dioscorea alata* polyploid species by cytogenetic and microsatellite segregation analysis [Abstract]. *International Conference on Polyploidy, Hybridization and Biodiversity*, 17-20 Mai, Saint-Malo, France.
20. Némorin A., Abraham K., David J., Maledon E., Arnau G. (2009). Inheritance patterns of tetraploid *Dioscorea alata* varieties. *ISTRC: Tropical Roots and Tubers in a Changing Climate: Convenient Opportunity for the World, Proceedings of the 15th Triennial Symposium*, 2-6 Novembre, Lima, Pérou.
21. Arnau G., Maledon E., Némorin A. (2007). Genetic improvement of the greater yam *D. alata* through polyploidy breeding. In: *Abstracts of Tropical Root and Tuber Crops (ISTRC)*, 2-6 Novembre, Lima, Pérou.
22. Arnau G., Nemorin A., Maledon E., Lambert F. (2007). Morpho-agronomic, cytogenetic and molecular characterization of the CIRAD yam collection for their enhancement and utilization in genetic improvement programs. In *Abstracts of the 42nd Annual meeting of the Caribbean Food Crops Society*, 16-21 Septembre 2007, San José, Costa Rica.
23. Sheela M.N., Arnau G., Abraham K., Sreekumari M.T. (2006). Comparative analysis of allelic diversity of Indian greater yams. Poster presented to the *14th Triennial Symposium of the International Society for Tropical Roots Crops*, 20-26 Novembre 2006, Thiruvananthapuram, Kerala, Inde.
24. Arnau G., Maledon E., Bachand I., Abraham K. (2006). Production of interploid hybrids and molecular markers heterozygosity determination using microsatellite markers in the greater yam, *D. alata*: Importance for the genetic improvement of the greater yam. *14th Triennial Symposium of the International Society for Tropical Roots Crops*, 20-26 Novembre, Thiruvananthapuram, Kerala, Inde.
25. Arnau G., Maledon E. (2005). Genetic improvement of yam: what strategy and objectives? Poster presented to the *41st Annual meeting of the Caribbean Food Crops Society*, 10-16 Juillet 2005, Gosier, Guadeloupe, France.

26. Malapa, R., Arnau G., Noyer J.L., Lebot V. (2002). Genetic relationships between *Dioscorea alata L.* and related Melanesian endemic species *D. nummularia* Lam. as revealed by AFLP markers. *43rd Annual Meeting of the Society for Economic Botany*, 22-27 Juin, New York.
27. Arnau G., Lallemand, J., Bourgoin M. (2000). Are AFLP markers the best alternative for cultivar identification? *International Symposium Molecular Markers for Characterizing Genotypes and Identifying Cultivars in Horticulture (ISHS)*, 6-8 Mars, Montpellier.
28. Arnau G., Lallemand, J., Bourgoin M. (2000). ISSR versus AFLP for variety identification: The example of strawberry. *Plant & Animal Genome VIII*, 9-12 Janvier, San Diego, Californie.
29. Arnau G., Lallemand, J., Bourgoin M. (1999). Development of AFLP markers for the identification of strawberry (*Fragaria ananassa* L.) varieties. *Gatersleben Research Conference Plant Genomics*, 17-21 Juin, Scholb Meisdorf.
30. Arnau G., Bourgoin M. (1998). Applicability of amplified fragment length polymorphism markers for strawberry variety identification. *5th Meeting of the International Union for the Protection of New Varieties of Plants (UPOV)*. Biochemical and molecular techniques, 28-30 Septembre, Beltsville, USA.
31. This D., Teulat B., Arnau G., Borries C., Souyris I., Borel C., Ragot C., Leroy P., Khairallah M., Monneveux P., Charrier A. (1997). QTLs for drought tolerance in barley. *V International Plant & Genome Conference*, 12-16 Janvier, San Diego, Californie.
32. This D., Arnau G., Teulat B., Lewicki S., Borel C., Simmoneau T., Borries C., Souyris I., Monneveux P. (1996). Which genes could be targets for drought tolerance improvement in barley? In: *Proceedings of VII International Barley Genetics Symposium*, Saskatoon, Canada, pp. 70-76.
33. Arnau G., This D., Monneveux P. (1995). Agronomical, physiological and molecular responses to drought stress in different barley cultivars: trying to find tolerance markers. *International Congress on Integrated Studies on Drought Tolerance of Higher Plants*, 31 Août - 2 Septembre, Montpellier, France.

DISTINCTION : 3e prix des meilleures présentations orales au *First World Congress on Root and Tuber Crops* (WCRTC) à Nanning, Guangxi, Chine, du 18 au 22 janvier 2016.
Titre de la présentation : *"Progress and challenges in genetic improvement of yam (Dioscorea alata L.)"*
Cette distinction récompense les travaux de recherche et d'amélioration génétique menés sur *Dioscorea alata* en Guadeloupe.

EXPERTISE

Reviewer pour Theor and Applied Genetics, Journal of Applied Genetics, Genes et Diversity

RESPONSABILITE INSTITUTIONNELLE

2002-2020 : Responsable de l'équipe "Igname" en Guadeloupe ; en charge de la gestion des activités et de la supervision de l'équipe.

Depuis 2021 : Animation de la vie collective de l'équipe DEFI et contribution aux réflexions scientifiques et stratégiques de l'équipe.

3.Cinq publications significatives

1. Némorin, A., David, J., Maledon, E., Nudol, E., Dalon, J., & Arnau, G. (2013).
Microsatellite and flow cytometry analysis to understand the origin of Dioscorea alata polyploids. Annals of Botany, 112(5), 811-819.

Ces recherches nous ont permis de déterminer les mécanismes les plus probables à l'origine des polyploïdes naturels de *Dioscorea alata* grâce à la cytométrie en flux et à la ségrégation de marqueurs microsatellites.

2. Arnau, G., Bhattacharjee, R., Chair, H., Malapa, R., Lebot, V., Abraham, K., Perrier, X., Petro, D., Penet, L., & Pavis, C. (2017).
Understanding the genetic diversity and population structure of yam (Dioscorea alata L.) using microsatellite markers. PloS One, 12(3), e0174150.

Il s'agit de la première étude mondiale consacrée à l'analyse de la diversité génétique de *Dioscorea alata*. Elle constitue une avancée majeure dans la compréhension de la diversité globale de cette espèce, en fournissant des données essentielles pour sa conservation et pour les travaux d'amélioration génétique.

3. Cormier, F., Lawac, F., Maledon, E., Gravillon, M. C., Nudol, E., Mournet, P., Vignes, H., Chair, H., & Arnau, G. (2019).
A reference high-density genetic map of greater yam (Dioscorea alata L.). Theoretical and Applied Genetics, 132(6), 1733-1744.

Nous avons généré la première carte génétique de haute densité pour *Dioscorea alata*, ce qui nous a permis d'identifier des SNPs associés aux divers traits analysés.

4. Ehounou, A. E., Cornet, D., Desfontaines, L., Marie-Magdeleine, C., Maledon, E., Nudol, E., Beurier, G., Rouan, L., Brat, P., Lechaudel, M., Nous, C., N'Guetta, A. S. P., Kouakou, A. M., & Arnau, G. (2021).
Predicting quality, texture and chemical content of yam (Dioscorea alata L.) tubers using near-infrared spectroscopy. Journal of Near Infrared Spectroscopy, 29(3), 128-139.

Cette étude présente le développement de la spectroscopie proche infrarouge (NIRS) comme méthode rapide et peu coûteuse pour prédire le contenu chimique des tubercules de *Dioscorea alata*. Cette technologie a permis de phénotyper efficacement nos populations, facilitant l'identification de QTLs liés à des traits physico-chimiques.

5. Arnau, G., Desfontaines, L., Ehounou, A. E., Marie-Magdeleine, C., Kouakou, A. M., Leinster, J., Nudol, E., Maledon, E., & Chair, H. (2024).
Quantitative trait loci and candidate genes for physico-chemical traits related to tuber quality in greater yam (Dioscorea alata L.). Journal of the Science of Food and Agriculture, 104(8), 4872-4879.

Cette étude nous a permis d'identifier des QTLs et des gènes candidats associés à quatre traits physico-chimiques clés influençant la qualité des tubercules de *Dioscorea alata*.

4. Synthèse des travaux de recherche

Axe 1 : Structure et fonctionnement du génome de l'igname *Dioscorea alata*

1. Acquisition de connaissances sur le nombre chromosomique de base et les niveaux de ploïdie

Lorsque j'ai commencé mes recherches au CIRAD en 2001, il était généralement admis que l'igname *Dioscorea alata* avait un nombre de chromosomes de base de x = 10, une croyance basée sur des travaux de recherche antérieurs, notamment ceux d'Essad et Maunoury (1984) et d'Abraham et Nair (1991). *D. alata* est une espèce polyploïde comprenant des accessions présentant trois niveaux de ploïdie distincts (2n = 40, 60, 80). À l'époque, on présumait que ces accessions étaient tétraploïdes, hexaploïdes et octoploïdes respectivement, en se basant sur l'hypothèse d'un nombre de chromosomes de base de x = 10.

En 2005, des études de ségrégation de marqueurs microsatellites ont révélé que l'espèce *D. rotundata*, autrefois considérée comme tétraploïde (2n = 40), était en réalité diploïde (Scarcelli *et al.,* 2005). Peu après, il a été démontré que l'espèce *D. trifida*, précédemment considérée comme octoploïde (2n = 80), était en fait tétraploïde, avec un nombre de chromosomes de base de x = 20 (Bousalem *et al*., 2006).

Dans le but de réévaluer le niveau de ploïdie de *D. alata* et d'obtenir une compréhension précise de son nombre de chromosomes de base, ces recherches ont été entreprises. Cette information revêt une importance cruciale pour optimiser l'efficacité des programmes d'amélioration génétique. Notre étude s'est concentrée sur l'analyse de la ségrégation de six marqueurs microsatellites au sein de quatre populations générées par croisements entre des génotypes ayant 40 chromosomes (Arnau *et al*., 2009).

Pour atteindre cet objectif, nous avons utilisé une approche bayésienne visant à explorer divers modes d'hérédité (diploïdie, allotétraploïdie et autotétraploïdie) et à déterminer le scénario de ségrégation le plus probable. Ces calculs prennent en compte toutes les hypothèses possibles concernant la configuration des allèles chez les deux parents, en relation avec les allèles observés dans leur descendance. À titre d'illustration, le tableau 1 synthétise la répartition des génotypes observés et attendus pour les différentes hypothèses de ségrégation testées dans le cas d'un locus comportant 3 allèles. La probabilité (ou vraisemblance) d'observer une distribution donnée a été déterminée en utilisant la distribution multinomiale, telle que décrite par Olson (1997). Cette méthode est pertinente pour comparer de nombreuses hypothèses et peut être utilisée même si les génotypes observés sont inattendus dans le cadre d'un scénario, ou si la taille des échantillons est faible, par rapport aux statistiques du chi carré (x2) (Olson 1997; Bousalem *et al*. 2006).

Pour les hypothèses d'hérédité tétraploïde (autotétraploïdie et allotétraploïdie), les proportions théoriques attendues de chaque génotype au sein des descendances, ainsi que les valeurs de vraisemblance, ont été calculées à l'aide d'un programme PERL développé par le Prof. Jacques David (Bousalem *et al*., 2006).

Tableau 1 : Distribution des génotypes attendus et observés dans la descendance (♀ "27F " X ♂ "24M") au locus Da2F10 sous hérédité disomique et tétrasomique.

Parental genotype	Progeny genotype						Likelihood (Log) P=0,01
	a	b	bc	ac	ab	abc	
Observed genotypes							
		6	17	13	24		
Distribution of expected progeny under disomic inheritance (2x, 1/4):							
ab x bc		1	1	1	1		-6
Distribution of expected progeny under tetrasomic inheritance (4x, 1/36):							
abbb x bccc			18			18	-77
aabb x bccc			6	3		27	-24
aaab x bccc				9		27	-90
abbb x bbcc		3	15		3	15	-34
aabb x bbcc		1	5	1	5	24	-34
aaab x bbcc				3	6	27	-53
abbb x bbbc		9	9		9	9	-24
aabb x bbbc		3	3		15	15	-39
aaab x bbbc					18	18	-63
Distribution of expected progeny under disomic inheritance (4x, 1/16):							
aa/bb x cc/bb*						16	-120
bb/aa x bb/cc*						16	-120
aa/bb x bb/cc						16	-120
aa/bb x bc/bc					4	12	-70
bb/aa x bc/bc					4	12	-70
bb/aa x cc/bb						16	-120
ab/ab x bb/cc			4			12	-82
ab/ab x bc/bc		1	3	1	3	8	-22
ab/ab x cc/bb			4			12	-82
ab/bb x bc/cc			8			8	-77
ab/bb x cc/bc			8			8	-77
bb/ab x bc/cc			8			8	-77
bb/ab x cc/bc			8			8	-77
ab/bb x bb/cc			8			8	-77
ab/bb x cc/bb			8			8	-77
ab/bb x bc/bc		2	6		2	6	-30
bb/ab x bb/cc			8			8	-77
bb/ab x cc/bb			8			8	-77
bb/ab x bc/bc		2	6		2	6	-30
ab/bb x bb/bc		4	4		4	4	-24
ab/bb x bc/bb		4	4		4	4	-24
bb/ab x bb/bc		4	4		4	4	-24
bb/ab x bc/bb		4	4		4	4	-24
aa/bb x bc/cc						16	-120
aa/bb x cc/bc						16	-120
bb/aa x bc/cc						16	-120
bb/aa x cc/bc						16	-120
ab/ab x bc/cc			4	2		10	-57
ab/ab x cc/bc			4	2		10	-57
aa/bb x bb/bc					8	8	-63
aa/bb x bc/bb					8	8	-63
bb/aa x bb/bc					8	8	-63
bb/aa x bc/bb					8	8	-63
ab/ab x bb/bc		2	2		6	6	-38
ab/ab x bc/bb		2	2		6	6	-38
aa/ab x bc/cc				4		12	-90
aa/ab x cc/bc				4		12	-90
ab/aa x bc/cc				4		12	-90
ab/aa x cc/bc				4		12	-90
aa/ab x bb/cc						16	-120
aa/ab x cc/bb						16	-120
aa/ab x bc/bc				2	4	10	-47
ab/aa x bb/cc						16	-120
ab/aa x cc/bb						16	-120
ab/aa x bc/bc				2	4	10	-47
aa/ab x bb/bc					8	8	-63
aa/ab x bc/bb					8	8	-63
ab/aa x bb/bc					8	8	-63
ab/aa x bc/bb					8	8	-63
Bayes factor hypothesis A : analysis integrating all the possibilities							6 e+16
Bayes factor hypothesis B : analysis excluding the possibilities to have the same allele in two differents locus. The considered alleles are marqued by an *							5 e+18

Un facteur de Bayes en faveur de la diploïdie par rapport à la tétraploïdie a été calculé pour chaque marqueur microsatellite, sous la forme d'un rapport entre les valeurs de probabilité de l'hypothèse de diploïdie (unique dans ce cas) et la somme des différentes hypothèses de tétraploïdie (Tableau 2). Les résultats obtenus ont clairement démontré que la ségrégation diploïde est bien plus probable que la ségrégation tétraploïde, une observation qui est particulièrement marquée pour les loci les plus discriminants, ceux comportant trois (abc) et quatre (abcd) allèles en ségrégation (400 < BF < 3E+20). En outre, les valeurs de l'indice de Bayes sont plus élevées en l'absence de phénomène d'homoplasie potentiel, c'est-à-dire lorsqu'il n'y a pas de possibilité d'avoir deux allèles de taille similaire dans deux loci indépendants.

Tableau 2 : Facteurs de Bayes testant la probabilité d'un modèle d'hérédité 2x par rapport à un modèle d'hérédité 4x dans les quatre populations en ségrégation. Les valeurs en gras indiquent que 2x est plus probable que 4x. L'analyse A intègre toutes les possibilités, tandis que l'analyse B exclue la possibilité d'avoir les mêmes allèles dans deux locus différents. BF>200 forte evidence (Arnau *et al.*, 2009).

Population	**Locus**	**Parental phenotypes**	**Total number of different allele**	**Bayes factors**	
				Hypothesis	
				A	**B**
24M x 27F (*N*=60)	Da2F10	ab x bc	3	**6×10^{16}**	**5×10^{18}**
	Dab2E07	ab x b	2	**1**	**1**
	Da1F08	a x ab	2	**1**	**1**
17M x St Vincent (*N*=23)	Da2F10	bd x ac	4	**7×10^{13}**	**3×10^{20}**
	Da3G04	ab x bc	3	**4×10^{9}**	**4×10^{16}**
	Dab2D11	ab x b	2	**3×10^{2}**	**3×10^{4}**
	Dab2D08	a x ab	2	**1**	**1**
	Da1F08	a x ab	2	0,1	**1**
31M x 27F (*N*=17)	Da2F10	bc x ab	3	**4×10^{2}**	**2×10^{4}**
	Dab2E07	ab x b	2	**1**	**1**
	Da1F08	a x ab	2	**4**	**4**
Pyramide x 27F (*N*=10)	Da2F10	bc x ab	3	**6×10^{2}**	**6×10^{2}**
	Dab2D11	ab x ac	3	**1×10^{3}**	**2×10^{5}**
	Dab2E07	ab x b	2	**4**	**20**
	Da1F08	a x ab	2	**1**	**1**
	Da3G04	ab x b	2	0,4	**2**

Par ailleurs, le nombre de chromosomes des accessions de notre collection de travail, a été évalué à l'aide de la cytométrie en flux ou par des comptages chromosomiques (Arnau *et al.*, 2009). Le dénombrement des chromosomes a été effectué sur les extrémités des racines mitotiques de quatre accessions (639a, 613a, 760a et 765a), connues pour présenter des nombres de chromosomes distincts (Malapa *et al.*, 2005), comprenant deux diploïdes (2n=40), un triploïde (2n=60) et un tétraploïde (2n=80). Ces quatre accessions ont servi de référence pour déterminer le nombre de chromosomes des autres accessions en utilisant la cytométrie en flux. La Figure 1 illustre une cellule au stade métaphasique de la variété triploïde 760a (2n=60) utilisée comme référence interne. Pour les analyses de cytométrie en flux, des noyaux de l'accession à tester et de la référence interne sont extraits, mélangés et marqués avec de

l'iodure de propidium, une molécule qui émet de la fluorescence lorsqu'elle est stimulée par le laser de l'appareil. Chaque génotype génère un pic spécifique majeur de fluorescence en G0-1, représentant la quantité d'ADN. La ploïdie de l'échantillon à tester est déterminée en comparant sa position relative à celle de la référence interne. À titre d'exemple, la Figure 1 présente les profils obtenus pour un clone avec 2n=40 chromosomes (S40) et un autre avec 2n=80 chromosomes (S80), en utilisant le clone 760a (2n=60) comme référence interne (IT60). Par le biais de ces analyses, nous avons pu montrer que notre collection de travail comprend 84 accessions diploïdes, 8 triploïdes, et 18 tétraploïdes.

De plus, la taille du génome de *D. alata* a été estimée en la comparant à celle d'une variété de citron "Lisbon", dont la taille absolue du génome avait été précédemment mesurée à 0,786 pg/2C (Ollitrault et Jacquemond, 1994). Cette comparaison a permis d'estimer les tailles relatives en pg/2C des accessions de D. alata diploïdes, triploïdes et tétraploïdes à environ 0,93 ± 0,08 pg, 1,43 pg et 1,98 ± 0,05 pg, respectivement.

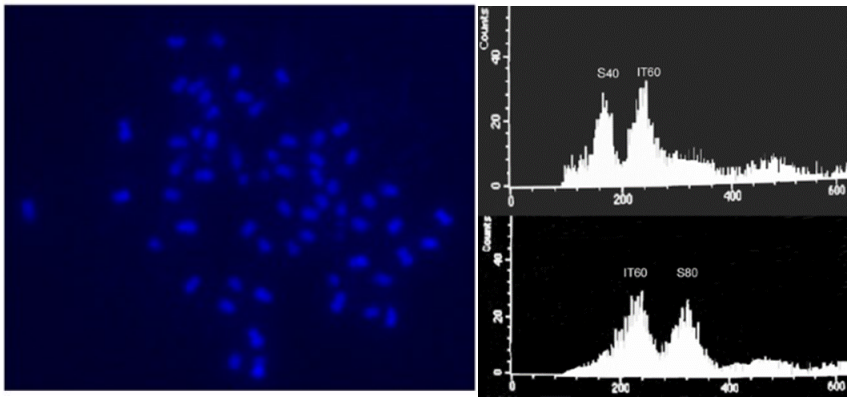


Figure 1: A, Comptage des chromosomes mitotiques à partir de préparations de pointes de racines de l'accession *D. alata* 760a (2n = 60). Les chromosomes ont été colorés au DAPI. B, Histogrammes de cytométrie en flux d'un clone avec 40 chromosomes (S40) et d'un clone avec 80 chromosomes (S80). S40 et S80 indiquent le pic G0-1 correspondant aux échantillons et IT60, le pic G0-1 correspondant à la référence interne ayant 60 chromosomes (Arnau *et al* ., 2009).

2. Analyse de l'hérédité des variétés tétraploïdes (auto-ou-allo-tétraploïdes)

Les premiers essais menés à la Station Expérimentale de Roujol en Guadeloupe avec notre collection de travail, ont mis en évidence le fait que les meilleures récoltes de tubercules (poids total par plante) étaient obtenues avec les variétés présentant les niveaux de ploïdie les plus élevés (Abraham et Arnau, 2007). De plus, ces essais ont révélé de la fertilité au sein des tétraploïdes naturelles originaires du Pacifique (2n=4x=80). Cette découverte a ouvert la voie à l'obtention des premières descendances polyploïdes, réalisées grâce à des croisements intercytotypes entre 2x X 4x ainsi que des croisements entre deux accessions tétraploïdes 4x X 4x (Abraham et Arnau, 2009 ; Arnau *et al.,* 2010).

Ces découvertes ont ouvert des perspectives prometteuses pour la production de variétés polyploïdes. Afin d'optimiser la création de matériels polyploïdes, il est essentiel de comprendre la nature génétique de ces derniers. En particulier, la compréhension de la

ségrégation de leurs chromosomes lors de la méiose, en relation avec leur constitution génétique, car cette caractéristique détermine le mode de transmission des allèles qu'ils contiennent et donc l'hérédité des caractères agronomiques.

Les plantes tétraploïdes peuvent avoir une origine allopolyploïde ou autopolyploïde. Dans le premier cas, elles résultent de la combinaison de deux génomes distincts et montrent une héritabilité de type disomique (association systématique des mêmes paires de chromosomes pendant la méïose). Dans le second cas, elles proviennent de la duplication d'un génome diploïde et montrent des appariements aléatoires entre les quatre chromosomes homologues dans chacune des classes d'homologie. La distinction entre accessions allo- et auto-tétraploïdes nécessite une étude de ségrégation chromosomique de leur descendance à l'aide de marqueurs spécifiques de locus, et/ou des études de cytogénétique.

Pour déterminer si les formes tétraploïdes sont d'origine auto- ou allotétraploïde, nous avons utilisé ces deux approches complémentaires. Une première étude a consisté à génotyper 20 marqueurs microsatellites (SSR) sur une descendance de 188 individus issue du croisement de deux géniteurs tétraploïdes naturels (♀ "Noulelcae" X ♂ "Tepuna"). Ces marqueurs microsatellites, qui ont révélé un maximum de 4 allèles chez chacun des deux parents, ont permis d'étudier la ségrégation chromosomique à des locus uniques dans le génome. Un test d'indépendance entre ces loci par une analyse deux points (logiciel TetraploidMap, Hackett et Luo 2003) a montré l'existence de liaisons génétiques entre 4 paires de marqueurs. Ces 20 marqueurs SSR représentent donc un ensemble de 12 marqueurs ségrégeant indépendamment.

Parmi ces 20 loci microsatellites, 3 d'entre eux ont révélé un ensemble de 5 allèles distincts dans la descendance, 9 loci ont révélé 4 allèles, et 8 loci ont révélé 3 allèles. L'analyse des ségrégations chromosomiques à chaque locus a été effectuée en utilisant la methode bayésienne (Olson, 1997), en comparant la vraisemblance statistique d'un modèle autotétraploïde à un modèle allotétraploïde. Les valeurs des facteurs bayésiens (BF) calculés pour chacun des 20 loci sont présentées dans le tableau 3.

Les facteurs bayésiens pour les 3 loci ségrégeant avec 5 allèles s'avèrent particulièrement élevés. Ces valeurs élevées montrent clairement qu'il existe une hérédité de type autotétraploïde au niveau de ces 3 loci. Parmi les 9 loci ségrégeant avec 4 allèles, 7 d'entre eux possèdent un BF supérieur à 200, confirmant également un mode de ségrégation tétrasomique (l'hypothèse autotétraploïde etant 200 fois plus probable que l'hypothèse alternative). Pour les deux autres loci restant à 4 allèles ainsi que les 8 autres à 3 allèles, il faut considérer dans l'hypothèse disomique un calcul de facteur bayésien prenant en compte la présence ou non d'éventuels phénomènes d'homoplasie (possibilité d'avoir deux allèles de même taille dans deux loci indépendants). En présence d'homoplasie, il n'y a pas de différence significative entre les vraisemblances disomique et tétrasomique. Toutefois, si on admet une absence d'homoplasie, l'hypothèse de l'hérédité tétrasomique l'emporte sur l'hypothèse alternative pour tous les 20 loci ($289 < BF < 1.33E+76$).

Tableau 3 : Facteurs de Bayes testant la probabilité d'une hérédité disomique versus polysomique 4x dans une population tétraploïde de *Dioscorea alata* en ségrégation sur 20 loci microsatellites. Nb A. : nombre d'allèles ; Σauto : somme de toutes les probabilités bayésiennes en cas où le mode de transmission est autotétrasomique ; Σallo homoplasie : somme de toutes les probabilités bayésiennes en cas de transmission disomique considérant l'homoplasie ; Σallo non homoplasie : somme de toutes les probabilités bayésiennes en cas de transmission disomique considérant l'absence d'homoplasie ; BF : facteur de Bayes,* BF>200 forte evidence (Nemorin *et al*., 2012).

Locus	Nb A.	Σauto	Σallo homoplasy	Σallo no homoplasy	BF homoplasy	BF no homoplasy
DIJ0443	5	3.37E-33	7.98E-76	7.98E-76	4.22E+42*	4.22E+42*
DIJ0342	5	3.06E-26	7.66E-64	7.66E-64	3.99E+37*	3.99E+37*
Dpr3B12	5	3.79E-17	5.65E-35	2.26E-73	6.71E+17*	1.68E+56*
Dab2D11	4	5.57E-11	2.38E-25	2.91E-33	2.34E+14*	1.91E+22*
mDaCIR8	4	3.54E-16	8.62E-26	1.00E-48	4.11E+09*	3.54E+32*
mDaCIR108	4	2.06E-19	4.48E-25	2.41E-45	4.60E+05*	8.55E+25*
mDaCIR51	4	1.37E-21	3.63E-27	5.39E-80	3.77E+05*	2.54E+58*
Dab2D07	4	3.73E-20	2.26E-24	9.30E-53	1.65E+04*	4.01E+32*
mDaCIR179	4	7.28E-13	4.47E-16	3.71E-79	1.63E+03*	1.96E+66*
mDaCIR66	4	1.26E-08	4.36E-11	4.36E-11	2.89E+02*	2.89E+02*
Dpr3F04	4	1.81E-06	1.14E-08	1.36E-82	1.59E+02	1.33E+76*
Da2F10	4	9.60E-18	5.15E-17	1.75E-111	1.86E-01	5.49E+93*
DIJ0012	3	1.30E-06	3.69E-08	1.64E-18	3.52E+01	7.93E+11*
Da3G04	3	2.93E-05	1.08E-06	1.38E-17	2.73E+01	2.12E+12*
Da1C12	3	1.42E-05	1.70E-06	1.82E-09	8.35E+00	7.80E+03*
Dab2E07	3	1.79E-06	3.75E-07	3.51E-12	4.77E+00	5.10E+05*
Dab2D08	3	4.90E-04	5.10E-04	1.33E-12	9.61E-01	3.68E+08*
mDrCIR81	3	9.00E-09	9.75E-09	2.21E-36	9.23E-01	4.07E+27*
Da1F08	3	7.56E-06	1.79E-05	2.13E-15	4.22E-01	3.55E+09*
Dpr3E10	3	2.19E-10	6.98E-09	6.04E-48	3.14E-02*	3.63E+37*

En conclusion, cette étude a permis de déterminer que les tétraploïdes sont d'origine autotétraploïde. Cette conclusion a été renforcée par l'observation de plusieurs tétravalents lors de la méiose du géniteur mâle (Figure 2), qui est une caractéristique typique des autopolyploïdes.

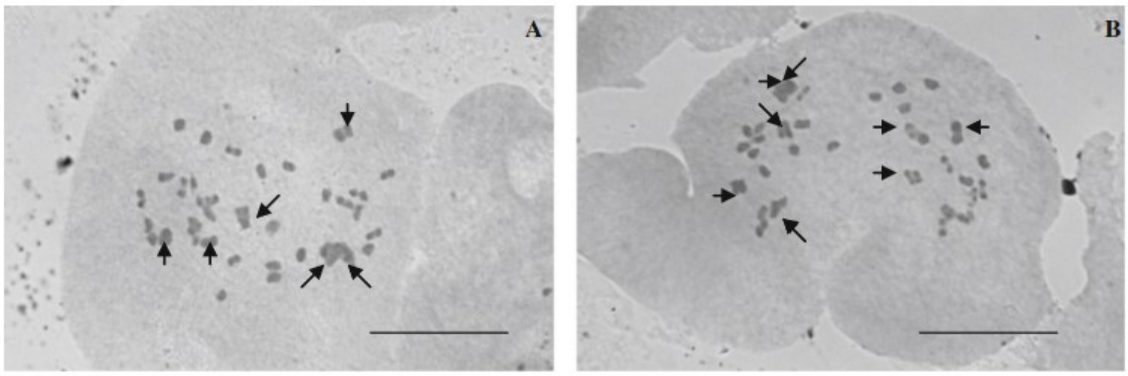


Figure 2 : PMC (Cellules mères de pollen) au stade métaphase I, montrant des quadrivalents et des bivalents. A, PMC montrant 6 quadrivalents et 28 bivalents, B, PMC montrant 8 quadrivalents et 24 bivalents (Abraham *et al*., 2013).

3. Etude sur l'origine des polyploïdes spontanés (triploïdes et tétraploïdes)

Différents croisements intracytotypes et intercytotypes ont été réalisés en Guadeloupe dans le but de déterminer les combinaisons de croisements viables permettant de produire des génotypes polyploïdes. Lors des croisements 2x X 2x, quelques individus triploïdes spontanés ont été obtenus (Arnau *et al*. 2006). De plus, les croisements intercytotypes ♀ 2x X ♂ 4x ont engendré des descendances triploïdes via le sauvetage d'embryons immatures (Figure 3), ainsi que quelques individus tétraploïdes (Arnau *et al.* 2006). En revanche, aucun résultat n'a été obtenu avec des croisements intercytotypes ♀ 4x X ♂ 2x, se traduisant par le gonflement de fruits sans graines. Enfin, des tentatives de croisements impliquant des femelles triploïdes et des mâles diploïdes ou tétraploïdes ont été faites, mais aucune descendance n'a été obtenue.

Ainsi, l'obtention de triploïdes au sein des descendances issues de croisements ♀ 2x X ♂ 2x et de tétraploïdes au sein des descendances issues de croisements ♀ 2x X ♂ 4x suggèrent que la formation de gamètes non réduits pourrait être à l'origine des phénomènes de polyploïdisation chez *D. alata*. Cependant, cette origine probable n'avait jusqu'alors été démontrée, et les mécanismes impliqués dans la formation des polyploïdes étaient encore à élucider.

L'objectif de cette étude était de comprendre l'origine des polyploïdes spontanés de *D. alata* en utilisant la cytométrie en flux et des marqueurs microsatellites. Pour parvenir à cette fin, différentes descendances ont été générées et analysées afin de mieux comprendre les phénomènes d'incompatibilité au niveau de l'albumen et de déterminer l'origine des gamètes en évaluant le taux d'hétérozygotie transmis à la descendance.

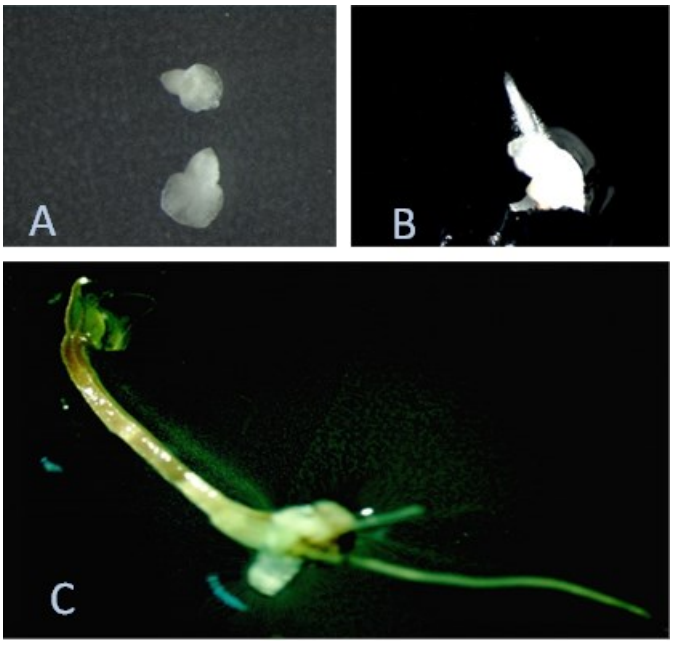


Figure 3 : Génération de plantules triploïdes par la culture in vitro d'embryons immatures (stade torpedo) issus de croisements intercytotypes (♀ 2x X ♂ 4x).

L'albumen, spécifique aux angiospermes, se développe après la double fertilisation, où un gamète mâle féconde l'oosphère pour former le zygote, tandis que l'autre gamète mâle féconde la cellule centrale, composée de deux noyaux polaires, pour former l'albumen. L'albumen est triploïde, avec un rapport de 2:1 entre le génome maternel et le génome paternel. Le développement normal de la graine in vivo est conditionné par l'albumen. Lors de croisements intercytotypes, un développement anormal de l'albumen entraîne fréquemment la non-viabilité des graines. La plupart des angiospermes ne produisent pas de descendances triploïdes viables lors de croisements intercytotypes, ce phénomène étant appelé le "bloc triploïde" (Marks 1966). Le ratio 2:1 entre le génome maternel et le génome paternel est crucial, car tout écart par rapport à ce ratio conduit à un albumen défectueux et à une graine non viable. Par exemple, lors de croisements 4x X 2x et 2x X 4x, le développement de l'albumen devient anormal en raison des rapports respectifs de 4:1 et 1:1 entre les génomes femelle et mâle. Cependant, si des gamètes 2n sont formés chez le parent diploïde, le ratio redevient 2:1, et le développement de l'albumen redevient normal. Ainsi, les barrières post-zygotiques liées aux incompatibilités au niveau de l'albumen peuvent être contournées grâce aux gamétophytes 2n (Carputo *et al*. 1997).

Une descendance de 2000 graines, issue du croisement entre le clone femelle diploïde et le clone mâle tétraploïde (♀ "5F" X ♂ "Tepuna") a été utilisée : (1) pour étudier l'incompatibilité de l'endosperme dans les croisements 2x X 4x ; et (2) pour analyser les événements de réduction non gamétique chez le parent femelle. Mille graines ont été semées pour évaluer la viabilité des graines et vérifier la ploïdie des plantules émergentes par cytométrie de flux. Les 1000 graines restantes ont été desséchées pour permettre une analyse séparée de la ploïdie de l'endosperme et des embryons. Sur les 1000 graines semées, seulement 18 plantules ont germé. La cytométrie en flux a montré que ces plantules étaient toutes tétraploïdes. Sur les 1000 graines desséchées, 995 avaient un endosperme anormal fripé (Figure 4) et seulement 5 avaient un endosperme normal. Les analyses de cytométrie en flux sur les endospermes

anormaux ont montré qu'ils étaient tétraploïdes, correspondant à l'endosperme 4x attendu si la femelle diploïde produisait un ovule haploide (n) et si l'endosperme provenait de la fusion des deux cellules n (2x) du sac embryonnaire avec le gamète mâle normal 2x produit par le mâle 4x. Ainsi, le ratio anormal observé 1:1 entre le génome maternel et le génome paternel explique que ce type de graines soit inviable, à moins que l'on réalise le sauvetage d'embryons par culture in vitro.

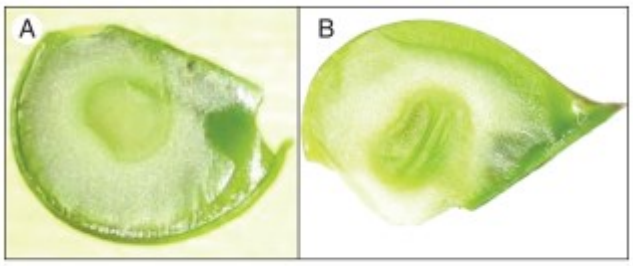


Figure 4 : Illustration d'une graine normale charnue (A) et une graine anormale fripée (B), mettant en évidence les caractéristiques distinctes entre les graines viables et non viables (Nemorin *et al*., 2013).

Une deuxième descendance de 300 graines a été obtenue par le croisement de deux diploïdes (♀ "5F" X ♂ "Kabusa"). Ces deux géniteurs sont soupçonnés de produire des gamètes non réduits. La moitié de ces graines (n = 150) a été semée, et des feuilles fraîches des plantules résultantes ont été soumises à une analyse par cytométrie de flux. L'autre moitié des graines a été soumise à une période de dessiccation de 90 jours après la pollinisation. À ce stade, les embryons ont été séparés de l'endosperme, puis mis en culture in vitro, comme le montre la figure 3. Des analyses de cytométrie de flux ont été réalisées à la fois sur les endospermes et sur les plantules issues de la culture in vitro des embryons. De plus, des analyses de ségrégation de marqueurs SSR ont été effectuées sur les feuilles des plantules identifiées comme triploïdes afin d''appréhender le taux d'hétérozygotie transmise par les parents à leur descendance.

Parmi les 300 descendants obtenus, quatre se sont avérés être triploïdes (1,3 %), tandis que 296 étaient diploïdes (98,7 %). L'apparence extérieure de toutes les graines semées ayant donné naissance à des plantules, qu'elles soient diploïdes ou triploïdes, était normale. Aucune différence morphologique significative n'a permis de distinguer a priori les graines triploïdes des graines diploïdes. L'analyse de la ploïdie des endospermes à l'origine des plantules triploïdes a pu être réalisée sur seulement deux des quatre plantules. Les résultats obtenus ont montré que les endospermes étaient triploïdes. De plus, l'étude de la transmission de l'hétérozygotie a permis d'exclure la polyspermie et la polyembryonie en tant que mécanismes à l'origine de ces triploïdes. En effet, cette analyse a révélé que les triploïdes obtenus avaient hérité soit de l'hétérozygotie de la mère, soit de celle du père. On peut donc en déduire qu'ils proviennent d'un ovule ou d'un grain de pollen non réduits.

En conclusion, nos résultats démontrent que la production de gamètes 2n est à l'origine des polyploïdes de D. *alata*, tout comme c'est le cas chez la pomme de terre (Iwanaga et Peloquin, 1982; Carputo *et al*., 2000), ainsi que chez de nombreuses autres espèces (Ramsey et Schemske, 1998).

Axe 2 : Diversité génétique des principales collections mondiales d'igname *D. alata*

Différents organismes nationaux et internationaux détiennent des collections importantes appartenant à l'espèce *D.alata* (l'INRA et le CIRAD en Guadeloupe, l'IITA au Nigeria et le CTCRI en Inde). La collection du CTCRI (Central Tuber Crops Research Institute, Inde) est constituée essentiellement de cultivars indiens. Celle de l'IITA (Institut international pour l'agriculture tropicale, Nigéria) de variétés provenant de différents pays d'Afrique de l'Ouest. La collection INRA-CRB-PT contient des variétés provenant d'origines géographiques diverses (Caraïbes, Pacifique Sud, Amérique du Sud), tandis que la collection de travail du CIRAD est principalement composée de génotypes provenant de Vanuatu (Pacifique Sud). Tous ces instituts mènent des travaux d'amélioration sur cette espèce.

L'objectif de cette étude était de caractériser les différentes collections mondiales avec les mêmes marqueurs afin de pouvoir comprendre les relations entre les différents pools génétiques. Le financement de cette étude a été assuré par une subvention de l'INRA (AIP Bio-Ressources : Div-Yam-2010) et des fonds FEDER Guadeloupe (Projets Caramba et Valexbiotrop).

Le génotypage des variétés (384 accessions dont 82 du CTCRI, 83 du CIRAD, 90 de l'IITA et 129 de l'INRA) a été réalisé à la Plateforme de Génotypage de Clermont Ferrand à l'aide d'un séquenceur capillaire. La lecture des allèles a été réalisée en Guadeloupe (Figure 5).

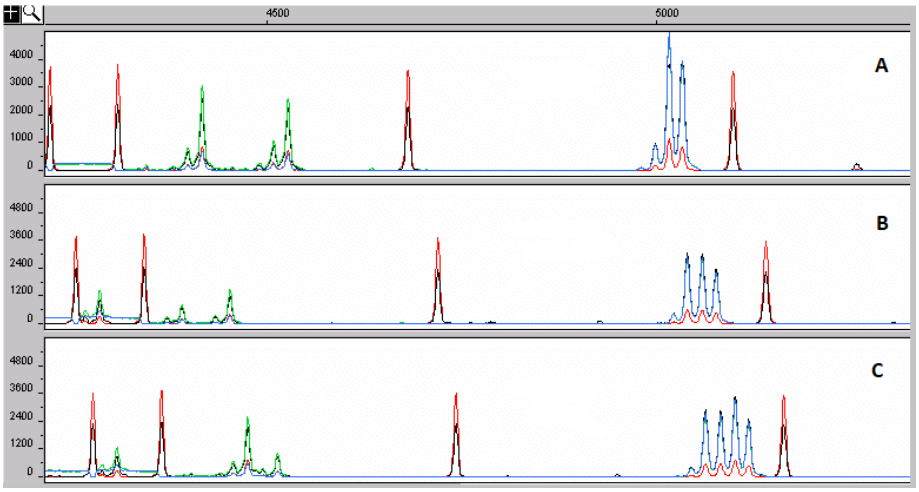


Figure 5: Exemples d'électrophorégrammes générés par un séquenceur d'ADN capillaire avec multiplexage d'échantillons. (A) Variété diploïde (2n=2x=40), (B) Variété triploïde (2n=3x=60), (C) Variété tétraploïde (2n=4x=80).

Les 384 variétés ont été évaluées à l'aide de 24 SSRs, ce qui a généré 311 allèles différents. Les paramètres suivants ont été notés pour chacune des collections: nombre total d'allèles observés (A^t), nombre moyen d'allèles par locus (A^m), nombre d'allèles spécifique à un groupe (A^s). La collection du CIRAD se distingue par un nombre total d'allèles et un nombre moyen d'allèles par locus plus élevé que les trois autres collections. Cette collection présente également un nombre d'allèles spécifiques (48) supérieur à ceux des autres collections (CTCRI : 14, IITA : 9, INRA : 10). Cette particularité peut s'expliquer par la présence plus importante de variétés polyploïdes.

Les données génétiques ont été transformées en une matrice binaire enregistrant la présence (1) ou l'absence (0) des allèles. Les distances génétiques entre chaque paire d'accessions ont

été calculées en utilisant les coefficients de dissimilarité de Dice, qui varient entre 0 et 0,86. Les individus présentant une distance de zéro partageaient les mêmes allèles sur l'ensemble des 24 loci et peuvent être considérés comme des doublons. Plusieurs doublons ont été identifiés dans chacune des collections. Un ensemble de six cultivars provenant de trois collections différentes a également montré des profils génotypiques identiques sur les 24 loci (tableau 4).

Tableau 4 : Détails sur les doublons au sein de chaque collection. Les accessions regroupées presentent des profils alléliques identiques sur les 24 loci SSR (Arnau *et al.*, 2017).

Collection	Accession code	Geographic origin	Local name	Study code
CTCRI	Da322	[4]India	Unknown	212
CIRAD	VU579	[3]Vanuatu	Letslets Bokis	318
CIRAD	VU567	[3]Vanuatu	Letslets Bolos	323
CRB-PT	PT-IG-00040	[2]Puerto Rico	59_Vino white forme	151
CRB-PT	PT-IG-00052	[2]Puerto Rico	71_Smooth Statia	168
CRB-PT	PT-IG-00395	Unknown	452_Fafadro bis	173
IITA	TDa-1427	[1]Ghana	Alamun Gaga	54
IITA	TDa-1437	[1]Ghana	Adidianmawoba	66
CTCRI	Da40	[4]India	Elivalan	192
CTCRI	Da73	[4]India	Muramchari	223
CTCRI	Da28	[4]India	Kachil	194
CTCRI	Da39	[4]India	Poolakachil	253
CTCRI	Da143	[4]India	Gutu	233
CTCRI	Da78	[4]India	Kachil	247
CTCRI	Da95	[4]India	Kudakachil	234
CTCRI	Da22	[4]India	Chuvanna Maveran	199
CTCRI	Da100	[4]India	Parisakodan	220
CTCRI	Da70	[4]India	Thekkan Kachil	222
CTCRI	Da120	[4]India	Kaduvakkayyan	228
CTCRI	Da105	[4]India	Chenithakizhangu	261
CTCRI	Da48	[4]India	Vila Kachil	255
CTCRI	Da209	[4]India	Kachil	243
CRB-PT	PT-IG-00061	[2]Martinique	80_Igname d eau	183
CRB-PT	PT-IG-00030	[2]Martinique	48_67	186
CRB-PT	PT-IG-00045	[2]Martinique	64_St Vincent Violet	370
CRB-PT	PT-IG-00067	[3]New Caledonia	86_Wénéféla bis	374

[1]Africa
[2]Caribbean
[3]South Pacific
[4]Asia

https://doi.org/10.1371/journal.pone.0174150.t002

La structuration de la diversité a été analysée à travers deux approches complémentaires : une analyse en coordonnées principales (ACP) sur l'ensemble des accessions et une analyse de regroupement de type cluster sur les accessions diploïdes en utilisant la méthode UPGMA. L'analyse combinée des distances de Dice et de l'analyse en coordonnées principales a permis d'identifier dix-sept groupes majeurs de cultivars étroitement apparentés, dont onze groupes de cultivars diploïdes, quatre groupes de triploïdes et deux groupes de tétraploïdes (Figures 6A, et 6B). Les distances génétiques au sein de ces groupes sont inférieures à 0,25.

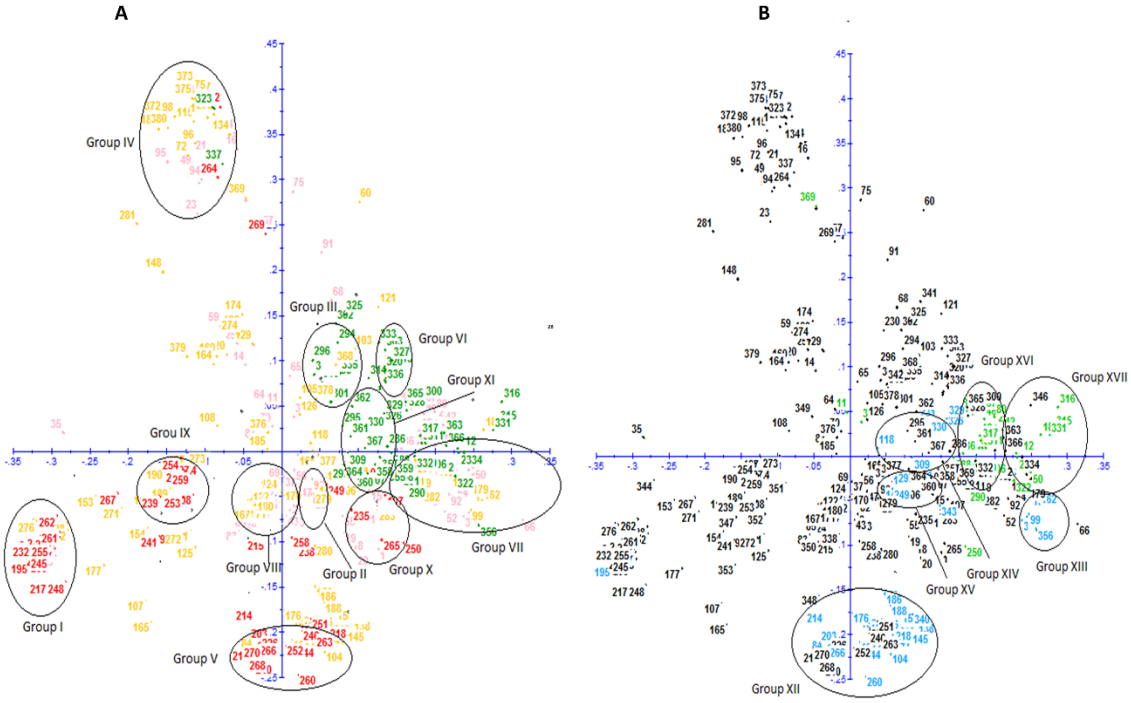


Figure 6 : A , Diagramme montrant les relations entre les 384 accessions de *D.alata* sur la base de l'analyse ACP. Les clones originaires du Vanuatu sont colorés en vert et ceux du CTCRI (Inde) en rouge. Les accessions du CRB-PT et de l'IITA sont colorées en orange et en rose, respectivement. **B**, Diagramme montrant les relations entre les 83 accessions polyploïdes basé sur l'analyse ACP. Les accessions tétraploïdes sont colorées en vert et les triploïdes en bleu (Arnau *et al*., 2017).

Cette étude a mis en évidence, l'existence d'une structuration qui est associée à l'origine géographique, aux niveaux de ploïdie et aux caractéristiques morpho-agronomiques. Les accessions originaires de l'Inde sont principalement distribuées dans la partie inférieure du graphique, tandis que les accessions du Vanuatu sont principalement distribuées dans la partie supérieure droite (Figure 6A), révélant deux pools génétiques distincts.

Les distances de Dice au sein de la collection indienne varient de 0 à 0,77, avec une valeur moyenne de 0,47, tandis que celles entre les cultivars du Vanuatu varient de 0 à 0,79, avec une moyenne de 0,49. Les distances génétiques entre les cultivars indiens et ceux du Vanuatu varient de 0,31 à 0,82, avec une valeur moyenne de 0,61. Cette valeur est significativement plus élevée que la valeur moyenne au sein de chacune des collections. De plus, certains allèles sont spécifiques à chacun de ces deux pools génétiques. Vingt-neuf allèles présents dans la collection indienne sont absents du pool génétique du Vanuatu, dont 12 allèles rares. De même, cinquante-cinq allèles présents dans le pool génétique du Vanuatu sont absents de la collection indienne, dont 15 allèles rares. Un test de Mantel effectué pour ces deux pools a démontré que la corrélation entre la proximité génétique et l'origine était significative ($r = 0,408$, $P < 0,0001$).

Les accessions des collections internationales (IITA et CRB-PT) sont réparties presque partout sur le l'ACP, ce qui montre que la diversité observée en Inde et au Vanuatu est bien représentée dans la collection du CRB-PT et partiellement représentée dans celle de l'IITA (Figure 6A). L'ACP a également montré que les cultivars originaires de Nouvelle-Calédonie (Pacifique Sud) sont principalement distribués dans la même partie du graphique (côté supérieur droit) que les cultivars du Vanuatu (Pacifique Sud).

Les 83 cultivars polyploïdes des quatre collections (49 triploïdes et 34 tétraploïdes) sont répartis dans différentes zones du graphique ACP et à proximité de cultivars diploïdes (Figure 6B). Cette répartition est en accord avec l'hypothèse selon laquelle les polyploïdes ont émergé par la formation de gamètes non réduits, issus des formes diploïdes.

L'analyse UPGMA a confirmé les résultats de l'ACP et a regroupé les diploïdes dans les mêmes onze groupes (Figure 7). Chaque groupe a été soutenu par des valeurs de bootstrap élevées (≥ 89 %), ce qui indique une grande stabilité des relations entre les accessions. Ces groupes reflètent donc une forte parenté phylogénétique et peuvent être considérés comme des regroupements naturels sur la base des valeurs bootstrap observées.

Les groupes I, IV et VIII sont ceux qui contiennent le plus grand nombre de cultivars (Figure 6A et Figure 7). A titre d'exemple, Le groupe IV rassemble 35 variétés i provenant de plusieurs régions géographiques (17 proviennent des Caraïbes, 11 de l'Afrique, 3 du Vanuatu, 2 de l'Inde, 1 de la Nouvelle-Calédonie et 1 de la Guyane française). Par ailleurs, ce groupe comprend six accessions présentant des profils alléliques identiques sur les 24 loci, ce qui les classe comme des probables doublons. Ces observations mettent en évidence la tendance des agriculteurs à adopter volontiers des cultivars présentant des caractéristiques supérieures, ce qui favorise leur diffusion à grande échelle. En résumé, nos résultats mettent en évidence une forte différenciation au sein de cette espèce, hautement heterozygote, probablement en raison de sa faible fertilité. Le mode de reproduction joue un rôle prépondérant dans la variabilité génétique, étant donné que les cultivars de *Dioscorea alata* se propagent essentiellement de manière végétative et présentent une faible fertilité. Deux hypothèses peuvent expliquer cette structuration en plusieurs groupes de cultivars génétiquement proches :

1. Les accessions appartenant à un même groupe pourraient résulter d'événements de reproduction sexuelle anciens impliquant les mêmes parents ou des cultivars génétiquement liés.
2. Les accessions appartenant à un même groupe pourraient provenir d'un clone initial identique qui a évolué par le biais de mutations somatiques fixées lors de la propagation végétative.

Étant donné le taux de mutation des marqueurs SSRs (compris entre 10E-3 et 10E-6), dépendant des espèces et de leur position dans le génome, et compte tenu des valeurs des distances génétiques observées au sein de ces groupes, il est probable que ces deux mécanismes ont contribué à la structuration de la diversité génétique de *D. alata*.

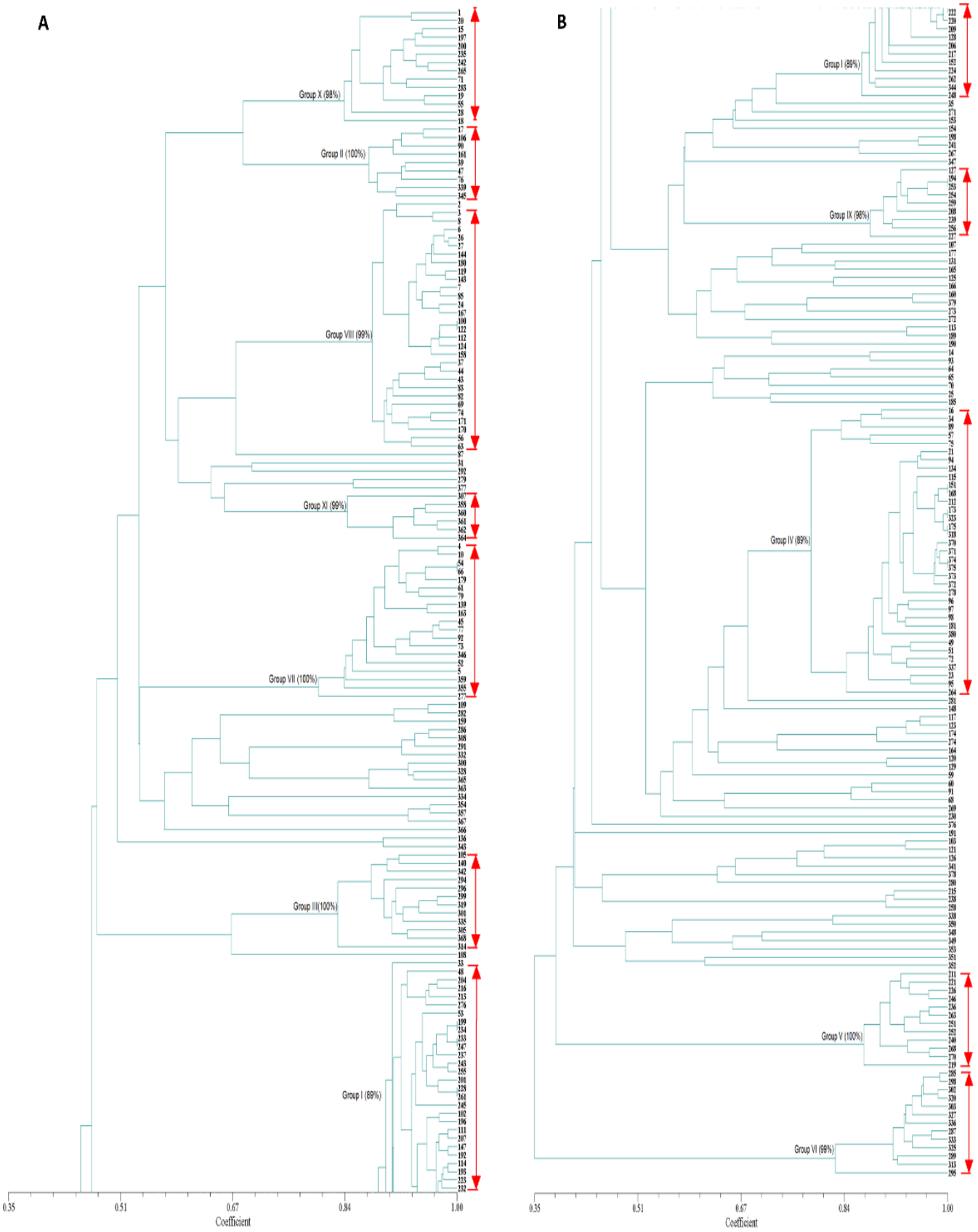


Figure 7 : Dendrogramme montrant les relations entre 284 accessions diploïdes de *D. alata* sur la base d'une analyse UPGMA avec 24 microsatellites**.**

Ce travail a également mis en lumière l'existence de deux pools génétiques distincts en Inde et au Vanuatu. Contrairement à l'espèce africaine *D. rotundata*, dont le centre d'origine est connu et la domestication bien documentée, le centre d'origine de *D. alata* demeure inconnu. Il est supposé que cette espèce a été domestiquée en Asie du Sud-Est il y a environ 6000 ans, avant de se propager en Inde et dans les îles du Pacifique Sud. Nos résultats, qui démontrent une nette différenciation génétique entre les cultivars de l'Inde et du Vanuatu, soutiennent l'hypothèse de l'existence de centres de diversification secondaire distincts en Asie et dans le Pacifique Sud.

En outre, cette étude a permis d'identifier divers groupes de cultivars florifères présentant une bonne résistance à l'anthracnose, ainsi que des ensembles renommés pour la qualité exceptionnelle de leurs tubercules. Les conclusions de cette recherche seront utiles pour les programmes d'amélioration en fournissant une aide dans la sélection de géniteurs génétiquement éloignés. Cette approche devrait contribuer à optimiser la diversité allélique au sein des descendances générées, favorisant ainsi une exploitation maximale de l'hétérosis.

Axe 3 : Exploration du déterminisme génétique de traits liés à la qualité des tubercules

La qualité des tubercules est d'une importance primordiale pour l'acceptation des nouvelles variétés. Elle repose sur une combinaison de divers traits morpho-agronomiques et de caractéristiques physico-chimiques qui influent sur leurs propriétés et leur texture. Lorsque nous avons entrepris nos travaux d'amélioration génétique sur *D.alata*, la base génétique des caractéristiques déterminant la qualité des tubercules était inconnue.

Ainsi, au début de notre programme d'amélioration, notre approche de sélection reposait essentiellement sur l'évaluation phénotypique des descendants afin d'identifier les hybrides les plus prometteurs. Nous nous concentrions sur plusieurs caractères connus pour influencer la qualité des tubercules, tels que la forme du tubercule, sa régularité, la couleur de la chair et le phénomène de brunissement de la chair après coupe (oxydation). Cependant, la sélection de certains de ces traits s'est avérée extrêmement difficile et chronophage, en particulier en ce qui concerne la régularité de la forme, car ce caractère est fortement influencé par les conditions environnementales, notamment la structure du sol.

1. Approche QTL (Quantitative Trait Loci) pour identifier des loci impliqués dans la variation de caractères morpho-agronomiques clés

Nous avons initié nos premières recherches visant à appréhender le déterminisme génétique des traits associés à la qualité du tubercule, dans le cadre d'un projet international intitulé "AfricaYam : Amélioration de la culture de l'igname pour augmenter la productivité et améliorer la qualité en Afrique de l'Ouest" (2015-2020), financé par la Fondation Bill et Melinda Gates. Ce projet a rassemblé 11 instituts de recherche de 8 pays différents, notamment le Nigeria (IITA, NRCRI, EBSU), la Côte d'Ivoire (CNRA), le Bénin (UAG), le Ghana (SARI, CRI), la Guadeloupe (CIRAD), le Japon (IBRC, JIRCAS), les États-Unis (Université Cornell) et l'Angleterre (JHI).

Les principaux objectifs de ce projet étaient les suivants : 1/ Établir une communauté d'améliorateurs dans les principaux pays producteurs d'Afrique de l'Ouest, en mettant l'accent

sur le perfectionnement des compétences en phénotypage, gestion des essais, gestion du programme, et méthodes de sélection. 2/ Mener des travaux de phénotypage pour des études d'association pangénomique et sur des populations biparentales, portant sur les traits agronomiques et de qualité essentiels.

Notre contribution à ce projet s'est principalement concentrée sur l'acquisition de connaissances sur le déterminisme génétique de plusieurs caractères morpho-agronomiques clés qui déterminent la qualité des tubercules (tableau 5). Pour atteindre cet objectif, nous avons adopté une approche de cartographie de QTLs (Quantitative trait loci) sur deux populations biparentales diploïdes F1 issues de croisements entre des géniteurs contrastés. Notre objectif était également d'identifier les régions du génome et des marqueurs moléculaires SNP associés à ces traits, dans la perspective de pouvoir mettre en place une sélection assistée par marqueurs pour faciliter le screening pour ces caractères.

Au début du projet, nous avons généré les deux populations ségrégeantes par pollinisations manuelles entre des géniteurs présentant des caractéristiques contrastées (♀ "74F" X ♂ "Kabusa" et ♀ "74F" X ♂ "14M"). Ces descendances proviennent de la même génitrice femelle et étaient composées de 193 et 121 hybrides respectivement. La femelle "74F" se caractérise par la production de tubercules longs, à la peau rugueuse, à la chair jaune, et présentant des racines tuberculaires (Figure 8). En revanche, les mâles "Kabusa" et "14M" produisent des tubercules cylindriques et ovales à la peau lisse, avec une chair blanche et blanc-crème. "Kabusa" ne développe pas de racines sur le tubercule, contrairement à "14M". Parmi les géniteurs, seule la femelle "74F" présente un brunissement oxydatif de la chair (Fig 8).

Tableau 5: Caractères morpho-agronomiques étudiés sur les populations A (♀"74F" X ♂ "Kabusa") et B (♀ "74F" X ♂ "14M") et les géniteurs.

Caractères	Modalités
Forme du tubercule (ratio L/W)	L/W (Longueur (L) and Largeur (W))
Regularité de la forme du tubercule	Élevé= 0 ; Faible= 1 ; Absence= 2
Racines tuberculaires	Absence= 0 ; peu= 1 ; beaucoup = 2
Texture de la peau	Lisse = 0 ; rugueuse= 1
Couleur de la chair (centre)	Blanc = 1 ; autres blancs = 2 ; autres couleurs=3
Oxydation de la chair (centre)	Non = 0 ; Oui= 1
Ratio d'oxydation	L0= longueur de l'oxydation 15 min après la coupe du tubercule frais

Figure 8 : Tubercules de la génitrice 74F (à gauche) et des mâles Kabusa (en haut à droite) et 14M (en bas à droite).

Pour accélérer leur multiplication, les deux populations en ségrégation ont été introduites in vitro par le biais du sauvetage d'embryons, puis multipliées par culture in vitro. De cette manière, tant les descendances que les géniteurs ont pu être plantés sur le terrain pendant deux années consécutives, en 2017 et 2018, en suivant un plan en blocs complets randomisés généralisés. Chaque génotype a été répliqué 18 fois (2 blocs x 9 plantes). Le phénotypage des descendances a été effectué sur une période de deux années consécutives, en 2018 et 2019. Les traits analysés et les modalités de mesure utilisées sont détaillés dans le tableau 5. Les observations concernant les quatre premiers caractères ont été réalisées sur la totalité des tubercules de chaque génotype (317 génotypes x 9 répétitions x 2 blocs x 2 années), tandis que les autres caractères ont été mesurés sur quatre tubercules de chaque bloc après avoir coupé les tubercules longitudinalement. La couleur de la chair a été notée immédiatement après la découpe du tubercule, tandis que le brunissement de la chair (oxydation) a été mesuré 15 minutes après coupe.

Une remarquable diversité a été constatée dans les deux populations en ségrégation pour la majorité des caractères, en particulier en ce qui concerne le ratio longueur/largeur et la régularité de la forme (Figure 9 et Figure 10). La plage de valeurs observées dans la population B, en ce qui concerne la régularité de la forme et la présence de racines tuberculaires, s'est avérée plus étendue que celle de la descendance A (Figure 10). L'analyse de la variance a révélé des effets significatifs du génotype et de l'année_bloc pour la plupart des caractères (tableau 6). Les analyses de variance et les héritabilités (tableau 6) ont mis en évidence l'importance des facteurs génétiques dans l'expression des caractères étudiés, présentant des valeurs d'héritabilité soit élevées, soit modérées. L'héritabilité de la régularité de la forme des tubercules a été signalée comme étant modérée (0,58) ou élevée (0,845) chez la pomme de terre, selon les études (Bradshaw *et al.,* 2008 ;Hara-Skrzypiec *et al*., 2018). Dans notre étude,

les héritabilités étaient significativement différentes entre les deux populations. L'héritabilité était modérée dans la population A (0,55),

Figure 9 : Variation phénotypique observée dans les descendances en ségrégation pour la forme des tubercules, la régularité de la forme, les racines tuberculaires et la couleur de la chair.

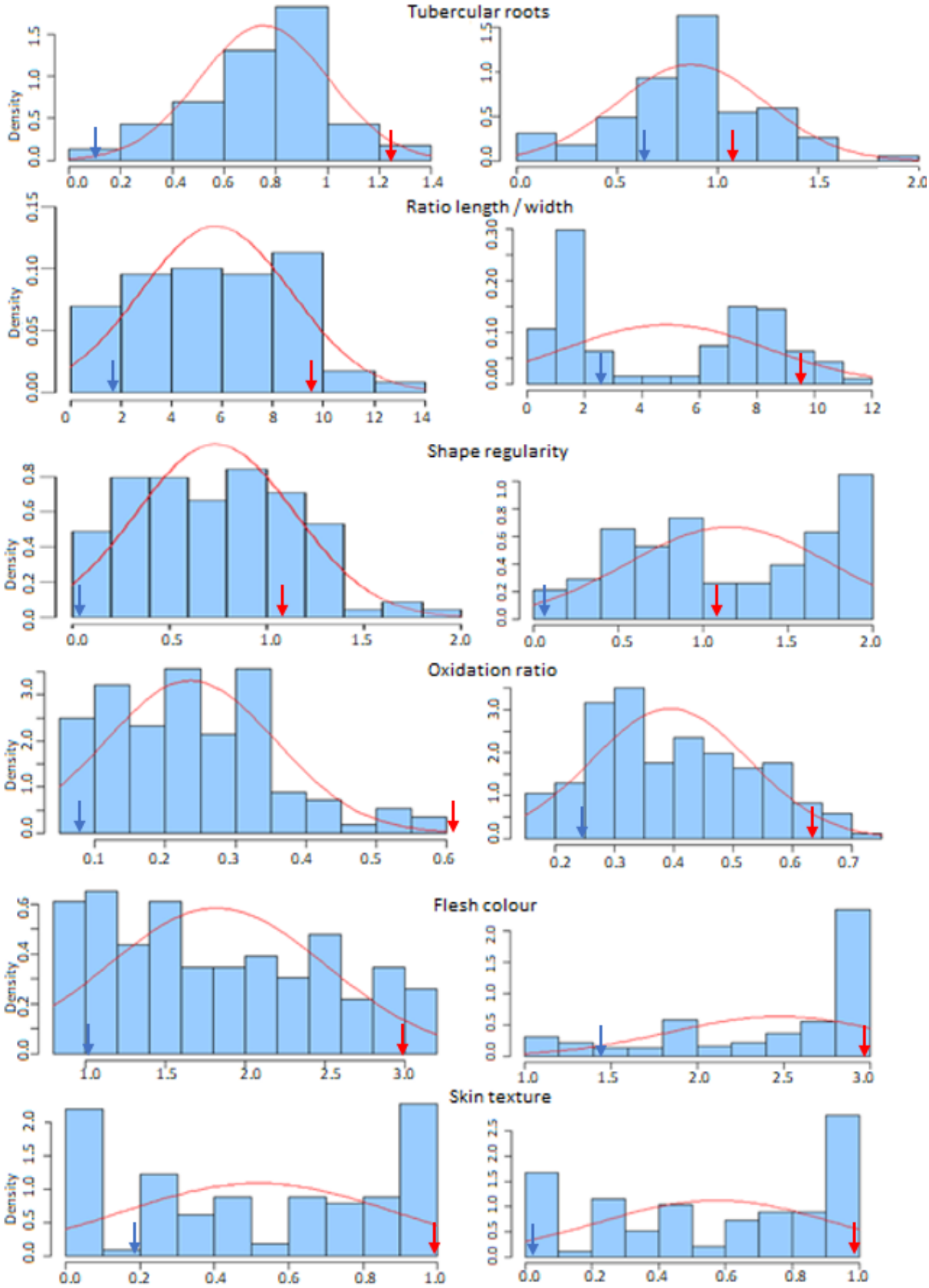


Figure 10 : Distribution des variations phénotypiques des caractères de qualité au sein de la population A (74F X Kabusa, à gauche) et de la population B (74F X 14M, à droite).

tandis qu'elle était élevée dans la population B (0,82). De même, une différence significative a été observée au niveau des racines tuberculaires, avec une héritabilité plus faible dans la population A (0,34) par rapport à la population B (0,51). Ces disparités peuvent s'expliquer par la plus grande variabilité observée au sein la population B pour ces deux caractères.

Tableau 6 : Analyses de variance du rapport longueur/largeur, des racines tuberculaires, de la texture de la peau, de la régularité de la forme, de la couleur de la chair, de l'oxydation de la chair et du taux d'oxydation (Ehounou *et al*., 2022).

	POPA					POPB				
Source	**df**	**Mean Sq**	**F**	***p***	**h^2**	**df**	**Mean Sq**	**F**	***p***	**h^2**
Ratio L/W										
Clone	111	30.90	12.04	*<0.0001*	0.92	180	38.25	13.81	*<0.0001*	0.94
Year_rep	3	14.91	5.81	*0.0007*		3	42.88	15.48	*0.0001*	
Tubercular roots										
Clone	111	0.21	1.48	*0.0052*	0.34	180	0.33	1.84	*<0.0001*	0.51
Year_rep	3	2.06	14.38	*<0.0001*		3	6.47	36.04	*<0.0001*	
Skin										
Clone	111	0.32	3.34	*<0.0001*	0.72	180	0.36	3.81	*<0.0001*	0.78
Year_rep	3	6.34	21.73	*<0.0001*		3	5.39	57.65	*<0.0001*	
Regularity of tuber shape										
Clone	111	0.51	2.02	*<0.0001*	0.55	178	1.07	4.55	*<0.0001*	0.82
Year_rep	3	0.35	1.37	0.2505 n.s		3	7.57	32.09	*<0.0001*	
Flesh colour										
Clone	111	5.30	17.57	*<0.0001*	0.96	175	3.56	16.04	*<0.0001*	0.96
Year_rep	3	23.77	78.80	*<0.0001*		3	14.66	66.09	*<0.0001*	
Flesh oxidation										
Clone	111	0.45	6.42	*<0.0001*	0.88	170	1.10	9.40	*<0.0001*	0.93
Year_rep	3	0.02	0.26	0.854 n.s		3	2.16	18.41	*<0.0001*	
Oxidation ratio										
Clone	110	0.07	9.40	*<0.0001*	0.91	159	0.11	6.42	*<0.0001*	0.87
Year_rep	3	0.10	6.04	*0.0004*		3	0.044	2.49	0.056 n.s	

Dans la population A, une distribution normale a été confirmée pour le rapport longueur/largeur, la régularité de la forme, les racines tuberculaires et le taux d'oxydation (p=0,05), tandis que les autres caractéristiques présentaient des distributions significativement déviées de la normalité. En revanche, dans la population B, les distributions étaient généralement asymétriques et/ou significativement éloignées de la normalité (p=0,05, Figure 10). Une ségrégation transgressive a été observée pour plusieurs caractères. Ce phénomène s'est manifesté de manière plus prononcée dans la population B, avec des valeurs significativement plus élevées que la génitrice femelle, notamment en ce qui concerne les racines tuberculaires et la difformité du tubercule.

L'émergence de ces hybrides transgressifs pourrait être attribuée à l'hétérozygotie des parents. Étant donné que les deux parents de cette population (74F et 14M) présentent un taux d'hétérozygotie très élevé (Cormier *et al.*, 2019), cela pourrait conduire à la création de nouvelles combinaisons alléliques au locus concerné, élargissant ainsi la gamme de variation phénotypique au sein de la population.

Nous avons mis en évidence des corrélations significatives entre plusieurs des caractères analysés dans les deux populations (Tableau 7). Dans l'ensemble, les corrélations se sont révélées plus prononcées dans la population B par rapport à la population A.

Tableau 7: Coefficients de corrélation (valeurs de Spearman) calculés entre le rapport longueur/largeur, la régularité de la forme, la texture de la peau, les racines tuberculaires, la couleur de la chair, l'oxydation de la chair, et le taux d'oxydation pour les populations A et B respectivement. n.s. : non significatif ; * : significatif à $p < 0,05$; ** : significatif à $p < 0,01$; *** : significatif à $p < 0,0001$.

Traits	Flesh colour	Ratio L/W	Skin texture	Tubercular roots	Regularity tuber shape	Flesh oxidation	Oxidation ratio
Flesh colour							
Ratio length/ width	0.423 n.s						
Skin texture	**0.586***	**0.694***					
Tubercular roots	0.028 n.s	−0.058 n.s	0.137 n.s				
Regularity tuber shape	0.378 n.s	0.089 n.s	0.410 n.s	0.169 n.s			
Flesh oxidation	**0.509***	0.411n.s	**0.417***	0.003 n.s	0.135 n.s		
Oxidation ratio	0.113 n.s	0.135 n.s	−0.067 n.s	−0.076 n.s	−0.186 n.s	0.284 n.s	

Traits	Flesh colour	Ratio L/W	Skin texture	Tubercular roots	Regularity tuber shape	Flesh oxidation	Oxidation ratio
Flesh colour							
Ratio length/width	**0.615***						
Skin texture	**0.590*****	**0.670****					
Tubercular roots	**−0.332*****	**−0.6241***	**−0.312***				
Regularity tuber shape	**−0.503*****	**−0.799*****	**−0.633****	**0.603*****			
Flesh oxidation	**0.644*****	**0.625***	**0.561***	**−0.296***	**−0.527***		
Oxidation ratio	**0.556*****	**0.621***	**0.590***	**−0.334***	**−0.569***	**0.761*****	

La cartographie des QTLs a été réalisée au moyen d'une carte consensus à haute densité, obtenue à partir de nos trois géniteurs par GBS (génotypage par séquençage, Cormier *et al.*, 2019). La recherche des loci impliqués dans la variation de chaque caractère a été menée sur chaque population en utilisant le logiciel MapQTL version 6 (Van Ooijen, 2009)

Ce travail a permis d'identifier des locus impliqués dans la variation phénotypique de chacun des caractères étudiés (Ehounou *et al*., 2022). Un plus grand nombre de QTLs a été détecté chez la population B que dans la population A (Tableau 8). Ce qui pourrait s'expliquer par la taille plus importante de la population B, ainsi que par une variabilité accrue observée au sein de cette population pour certains traits. La fraction de la variance phénotypique totale attribuable à un QTL spécifique variait entre 11,1 % et 43,5 %. Au total, 34 QTLs ont été détectés sur 8 des 20 chromosomes de l'igname (LG1, LG2, LG6-M, LG7, LG15, LG16, LG17 et LG19). Ils correspondent à cinq de chacun des caractères suivants : forme des tubercules, régularité de la forme, racines tuberculaires, texture de la peau, oxydation de la chair, six pour le taux d'oxydation et trois pour la couleur de la chair. Le nombre élevé de QTLs met en lumière la complexité des caractères étudiés. En revanche, un nombre moindre de QTLs a été détecté pour la couleur de la chair, suggérant qu'un nombre plus restreint de loci contrôlent ce caractère, ce qui est cohérent avec des observations antérieures sur la pomme de terre (Hara-Skrzypiec *et al*., 2018).

Plus de la moitié des QTs étaient positionnés sur le chromosome 16, ce qui indique que ce chromosome joue très probablement un rôle important dans le déterminisme génétique des caractères étudiés.

Nous avons identifié plusieurs colocalisations de QTL sur le chromosome LG16, impliquant un grand nombre de caractères (tableau 8, Figure 11). En outre, une colocalisation de QTL a été observée pour le rapport longueur/largeur des tubercules et la texture de la peau sur le chromosome LG15. Ces colocalisations étaient cohérentes avec les corrélations observées. Les caractères étudiés pourraient donc être sous le contrôle de plusieurs gènes distincts et étroitement liés, ou d'un seul gène ayant un effet pléiotropique. Les deux hypothèses semblent plausibles puisque les corrélations génétiques sont plus ou moins fortes selon les caractères. En effet, le rapport longueur/largeur était fortement et positivement corrélé avec la texture de la peau, ce qui suggère que ces deux caractères pourraient être contrôlés par les mêmes gènes. Chez la pomme de terre des corrélations génétiques significatives et des colocalisations de QTLs ont été trouvées également entre la forme du tubercule et la régularité de la forme, et entre la régularité de la forme et la couleur de la chair et entre la régularité de la forme et la couleur de la chair (Meijer *et al*., 2018; Hara-Skrzypiec *et al*., 2018).

Tableau 8 : QTLs (Quantitative trait loci) détectés pour les divers traits morpho-agronomiques étudiés au sein des deux populations en ségrégation A et B (Ehounou *et al.*, 2022).

Trait	LG	Marker associated	Marker origin	QTL position	LOD	R^2 (%)	Interval (%)
POPA							
Ratio L/W	LG16	16.1_2289706	M	25.1	4.51	17.5	0–32.6
	LG16	16.1_22654587	D	41.1	4.85	18.7	38.8–44.9
	LG15	15.1_6269438	D	80.0	3.61	14.3	71.6–87.7
	LG02	02.1_26794334	M	71.2	3.60	13.7	67.7–79.9
Skin texture	LG16	16.1_22761338	F	18.5	4.12	16.1	9.5–32.6
	LG16	16.1_22463422	D	44.1	5.88	22.2	39.7–44.9
	LG15	15.1_7629197	M	81.9	4.31	16.8	75.6–83.9
Flesh colour	LG16	16.1_23058008	M	16.1	5.24	20	4–28.1
	LG16	16.1_225392c67	F	41.0	4.94	19	38.8–44.9
Flesh oxidation	LG16	16.1_2248c9c68	F	14.3	5.68	21.5	6–22.3
	LG16	16.1_22648887	D	35.4	5.02	19.3	31.6–41.1
	LG15	15.1_8934406	M	123	4.72	17.5	122–125
	LG19	19.1_9513064	F	70.1	3.5	13.9	62.1–73.9
Oxidation ratio	LG16	16.1_22558120	F	22.4	12.28	40.8	12.4–70.4
	LG07	15.1_2205278	M	78.7	4.2	16.4	74.8–82.2
	LG06_M	06.1_31811266	M	57.8	3.57	14.1	51.5–62.3
POPB							
Ratio L/W	LG16	16.1_23043957	D	21.1	5.67	18.5	16.3–24.3
	LG16	16.1_22504963	D	38.9	6.23	20.1	38.8–49.5
	LG16	16.1_21653648	D	64.4	8.99	27.6	56.5–81.6
Regularity of tuber shape	LG16	16.1_23043957	D	20.7	6.25	19	18.5–24.1
	LG16	16.1_22504963	D	38.8	5.2	16	38.5–44.0
	LG16	16.1_21653648	D	64.4	11.37	32	58–67.4
	LG02	02.1_25284501	D	62.4	3.91	12.2	58.4–64.4
	LG01	01.1_400095	D	91.7	3.64	11.5	82.8–98.9
Tubercular roots	LG16	16.1_23043957	D	21.1	7.94	23.1	16.3–25.4
	LG16	16.1_22504963	D	38.8	7.77	22.7	38.8–45.5
	LG16	16.1_21653648	D	63.5	6.91	20.5	56.5–64.4
	LG19	21.1_1121998	D	163.1	4.39	13.5	156.8–163.1
	LG06_M	06.1_29183630	M	13.1	3.58	11.1	5.8–15.4
Skin texture	LG16	16.1_23043957	D	21.1	6.42	19.1	17.3–24.4
	LG16	16.1_2250496 3	D	38.8	5.7	18	38.0–44.1
	LG16	16.1_21653648	D	64.4	17.38	43.5	57.5–66.4
	LG17	17.1_7523202	M	86.3	4.25	13	82.6–94.5
Flesh colour	LG16	16.1_23043957	D	21.1	11.08	29.9	15.3–24.4
	LG16	16.1_22504963	D	38.9	9.36	28.1	38.5–49.0
	LG16	16.1_21653648	D	65.4	16.02	40	57.5–69.4
Flesh oxidation	LG16	16.1_23043957	D	21.1	5.7	18	15.3–24.4
	LG16	16.1_22504963	D	38.8	4.1	13.3	38.8–49.5
	LG16	16.1_21653648	D	64.4	7.04	21.1	55.5–81.6
Oxidation ratio	LG16	16.1_22558120	D	21.7	5.66	18.4	16.1–25.1
	LG16	16.1_22504963	D	38.8	6.23	20.1	38.5–39.7
	LG16	16.1_21653648	D	66.4	8.99	27.6	56.5–81.7
	LG07	07.1_1751053	D	103.1	3.67	12.4	102.8–107.8

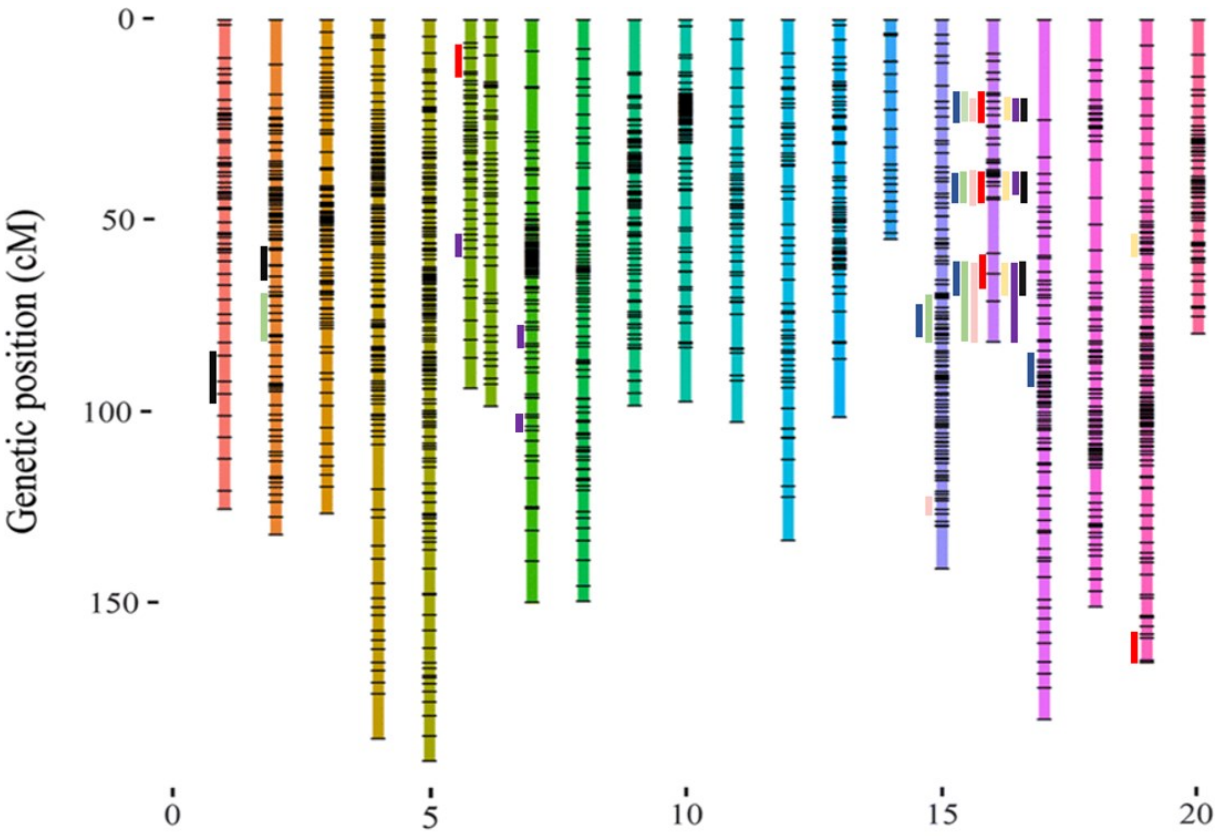


Figure 11 : Localisation sur la carte de référence des QTLs (Quantitative Trait Loci) identifiés pour les divers traits morpho-agronomiques étudiés : forme du tubercule (vert), régularité de la forme (noir), racines tuberculaires (rouge), texture de la peau (bleu), couleur de la chair (rose), oxydation de la chair (jaune) et le ratio d'oxydation (violet).

2. Cartographie de QTLs impactant la qualité physico-chimique des tubercules

Parallèlement, des investigations ont été entreprises pour analyser l'architecture génétique de plusieurs caractéristiques physico-chimiques qui déterminent la qualité organoleptique et texturale des tubercules. Ces travaux ont été réalisés dans le cadre d'un autre projet financé par la Fondation Bill et Melinda Gates, intitulé "RTBFoods: Breeding RTB products for end user preferences" (2018-2022). Les paramètres examinés comprennent la teneur en amidon, la matière sèche (DMC), les sucres et les protéines. Ces recherches ont été conduites en utilisant les mêmes populations biparentales que pour l'étude antérieure. Le phénotypage de ces diverses caractéristiques chimiques a été réalisé au moyen de la méthodologie NIRS, mise au point au début du projet en collaboration avec des qualiticiens de l'unité de recherche UR ASTRO de l'INRAE, en Guadeloupe (Ehounou *et al*., 2021)

L'ANOVA sur les données phénotypiques de la population A a montré des effets très significatifs ($p< 0,0001$) du génotype et des effets significatifs de la répétition ($P<0,05$) pour tous les caractères. Cependant, les génotypes étaient la source de variation la plus importante pour chaque caractère étudié. L'ANOVA des données phénotypiques de la population B a montré des effets génotypiques hautement significatifs ($p< 0,0001$) pour tous les caractères étudiés. Les héritabilités pour les quatre caractères variaient entre 0,68 à 0,88 dans la population B et

0,69 à 0,81 dans la population A . Les héritabilités pour l'amidon étaient similaires dans les deux populations (0,68 et 0,69) ainsi que pour les protéines (0,78 et 0,81). Celles obtenues pour le DMC et la teneur en sucre étaient significativement plus élevées dans la population B (0,86 et 0,88) que dans la population A (0,75 et 0,72).

Les distributions phénotypiques des quatre caractères étaient conformes à une distribution normale au sein des deux populations, comme illustré dans la Figure 12, reflétant des variations quantitatives typiques. Dans la population A (74F x Kabusa), les taux d'amidon présentaient une variation allant de 74,0 % à 83,6 %, tandis que dans la population B (74F x 14M), cette variation s'étendait de 71,1 % à 85,6 %. En ce qui concerne le DMC, la fourchette des scores était de 26,7 % à 38,8 % dans la population A et de 27,6 % à 41,3 % dans la population B. Les concentrations en sucres variaient de 0,58 % à 5,08 % dans la population A et de 0,10 % à 5,84 % dans la population B. Enfin, les taux de protéines présentaient une plage de variation de 3,95 % à 7,64 % dans la population A et de 4,05 % à 7,90 % dans la population B.

Concernant les géniteurs mâles, 14M et Kabusa, ils présentaient des valeurs significativement différentes de celles de la femelle 74F pour l'amidon (80,5 %, 79,2 %, 77,2 % respectivement), le DMC (31,5 %, 28,9 %, 28,1 %) et les sucres (1,79 %, 4,12 %, 3,22 %). Cependant, aucune différence significative n'a été observée entre les géniteurs en ce qui concerne la teneur en protéines. Des ségrégations transgressives ont été observées, avec des valeurs phénotypiques soit inférieures, soit supérieures à celles des parents pour tous les caractères (Figure 12).

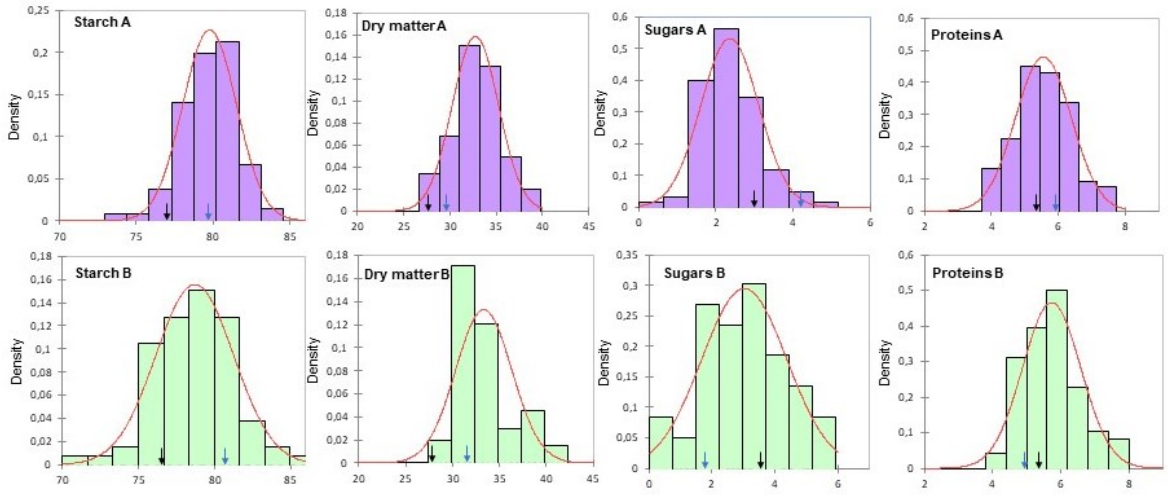


Figure 12 : Distributions phénotypiques pour l'amidon, la teneur en matière sèche, le sucre et la teneur en protéines dans les populations de cartographie A (74F × Kabusa) et B (74F × 14 M) (Arnau *et al.,* 2023).

Plusieurs corrélations significatives ont été détectées entre les caractères. Des corrélations négatives ont été trouvées au niveau des deux populations entre la teneur en amidon et la teneur en protéines, et entre la teneur en amidon et la teneur en sucre. Une corrélation positive a été détectée dans la population A entre la teneur en amidon et le DMC, tandis que dans la population B, une corrélation positive a été détectée entre le DMC et les sucres.

Un total de 25 QTL a été identifié pour les quatre caractères étudiés dans les deux populations de cartographie. Pour la teneur en amidon, quatre QTL ont été détectés dans la population B, situés sur les chromosomes 2, 5, 6 et 10, qui expliquent respectivement 22,4, 22,6, 20,1 et 22,1 % de la variance phénotypique (tableau 9, Figure 13). Dans la population A, trois QTL ont été détectés sur les chromosomes 10, 11 et 18, expliquant respectivement 15,4, 19,1 et 16,1% de la variance phénotypique totale. Les deux QTLs identifiés sur le chromosome 10 étaient situés dans des régions distinctes et sont deux loci différents. La figure 14 montre à titre d'exemple celui qui a été identifié dans la population B. Pour le DMC, quatre QTL ont été identifiés dans la population B, situés sur les chromosomes 1, 4, 7 et 12, qui expliquent 98,5 % de la variance phénotypique totale. Dans la population A, deux QTL ont été trouvés sur les chromosomes 1 et 2, expliquant respectivement 14,3 % et 18,7 % de la variance phénotypique. La figure 14 présente des informations sur le QTL identifié sur le chromosome 2 (population A). Trois QTL ont été détectés pour les sucres dans la population B (chromosomes 7, 9 et 13) et trois dans la population A (chromosomes 6, 7 et 12), expliquant 68,5% et 58,8% de la variance phénotypique totale dans chaque population, respectivement. La figure 14 montre les résultats du QTL trouvé dans le chromosome 7 (population A). Enfin, quatre QTL ont été révélés pour les protéines dans la population B (chromosomes 2, 5, 8 et 19) et deux dans la population A (chromosomes 10 et 18), expliquant 98,4 % et 38,4 % de la variance phénotypique totale de chaque population. La figure 14 montre les résultats du QTL trouvé sur le chromosome 19.

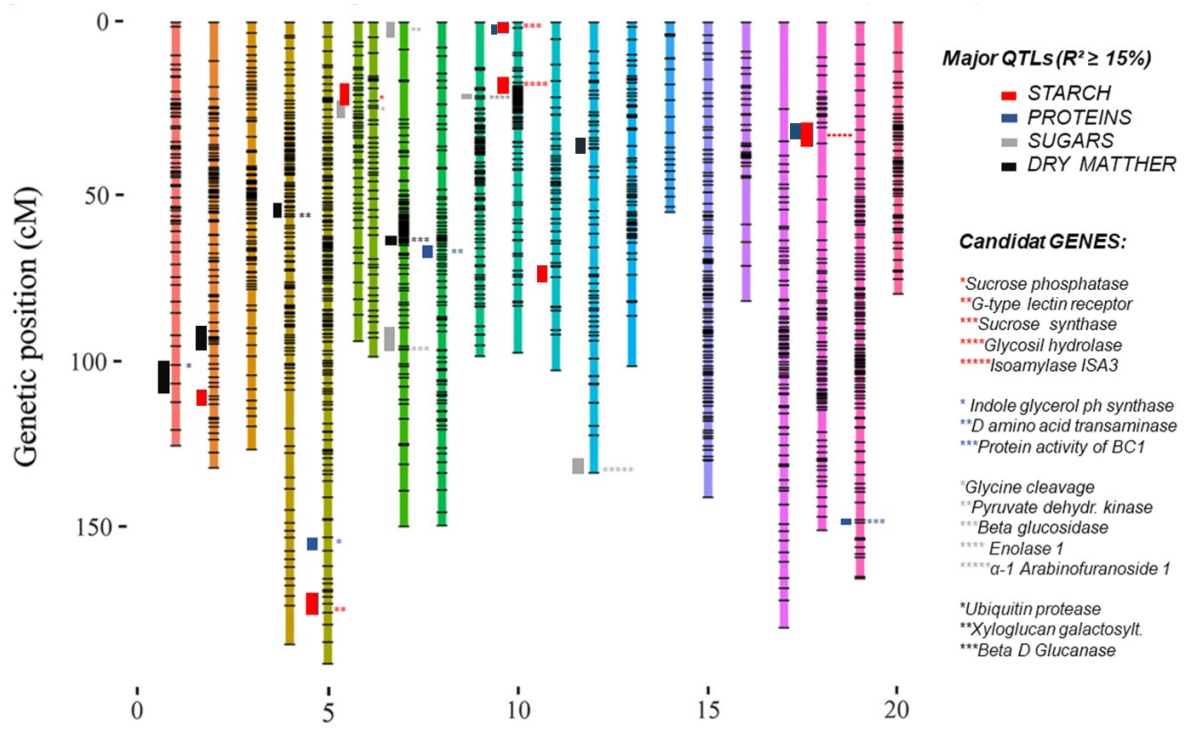


Figure 13 : Localisation sur la carte de référence des QTLs (Quantitative Trait Loci) majeurs ($R^2 > 15$ %) identifiés pour les quatre traits physico-chimiques étudiés. Les gènes candidats identifiés dans les intervalles de confiance des QTLs ont également été positionnés.

Tableau 9: Qtls validés pour les teneurs en amidon, DMC, sucres et protéines dans le panel de diversité et gènes candidats identifiés.

Trait	Chr	Pop.	LOD	R2 (%)	SNP	Alleles / Allele effect	Localisation	Candidate gene
Starch	5	B	3.67	22.6	05.1_32707706**	CC/CT/TT (T-)	Intergenic	G- type lectin Receptor like serine/threonine protein kinase
Starch	2	B	3.63	22.4	02.1_32141722*	TG/GG (G-)		
Starch	6	B	3.39	20.1	06.1_20308422*	AA/A*/**(*+)	Intergenic	Sucrose phosphatase 1 EC 3.1.3.24
Starch	10	B	3.58	22.1	10.1_6610171*	T*/**(*-)	Intergenic	Glycosil hydrolase family 5
Starch	10	A	3.24	15.4	10.1_17600318*	CC/C* (*+)	Intergenic	Sucrose phosphate synthase EC 2.4.1.14
Starch	11	A	4.10	19.1	11.1_4493720**	TT/TC (C-)	Intronic	Uncharacterized protein Loc120271825
Starch	18	A	3.40	16.1	18.1_1695612*	GG/GA/AA (A+)	Intronic	Isoamylase 3 cloroplastic EC 3.2.1.68
Proteins	8	B	4.46	26.6	08.1_18931363*	CC/CG (G-)	Intergenic	D amino acid transaminase EC2.6.1.21
Proteins	19	B	4.33	26.1	21.1_107691*	CC/CT/TT (T+)	Intergenic	Protein activity of BC1 complex kinase 7
Proteins	5	B	3.85	23.6	05.1_31811203*	TT/T* (*-)	Intergenic	Indole-3-Glycerol phosphate synthase EC 4.1.1.48
Proteins	2	B	3.54	21.9	-	non val.		
Proteins	18	A	4.90	22.4	18.1_1629151**	GG/GA/AA (A+)	Exon	Uncharacterized protein Loc120282302
Proteins	10	A	3.36	16.1	10.1_16586803**	GG/GC/CC (C-)	Intergenic	Putative proline rich receptor like protein kinase
Sugars	9	B	4.07	24.7	09.1_19344114**	A*/** (*-)	Intergenic	Enolase 1 chloroplastic EC 4.2.1.11
Sugars	7	B	3.91	23.9	07.1_351152*	GG/GA/AA (A-)	Intronic	Pyruvate dehydrogenase kinase EC 1.2.4.1
Sugars	13	B	3.18	19.9	-	non val.		
Sugars	12	A	4.60	21.2	12.1_17583120**	TG/GG (G+)	Intergenic	Alpha 1 Arabinofuranosidase 1 EC 3.2.1.55
Sugars	6	A	4.49	20.7	06.1_18785797	AA/A* /** (*-)	Intergenic	Glycine cleavage system H Protein 2
Sugars	7	A	3.51	16.6	07.1_1765193**	CT/TT (T+)	Intergenic	Beta glucosidase 22
DMC	4	B	4.84	27.2	04.1_10157871*	TT/TC/CC (C-)	Intergenic	Xyloglucan galactosyltransferase GT17
DMC	12	B	4.77	25.1	12.1_24347534*	TT/TC/CC(C-)		
DMC	7	B	4.13	24.0	07.1_3805155*	GG/GA (A+)	Intergenic	Endo 1,3 (4) beta D Glucanase
DMC	1	B	3.79	22.2	01.1_601071*	GG/**/** (*-)	Intronic	Ubiquitin like specific protease 2B
DMC	1	A	3.00	14.3	-	non val.		
DMC	2	A	3.99	18.7	02.1_28100114**	GA/AA (A-)		

Chr Chromosome, *Pop* Population, *SNP* marker in the QTL Confidence Interval that showed a significant association with phenotypic data from the diversity panel. **Significant at P < 00.1, * P<00.5. *Non val* QTLs non validated in the diversity panel

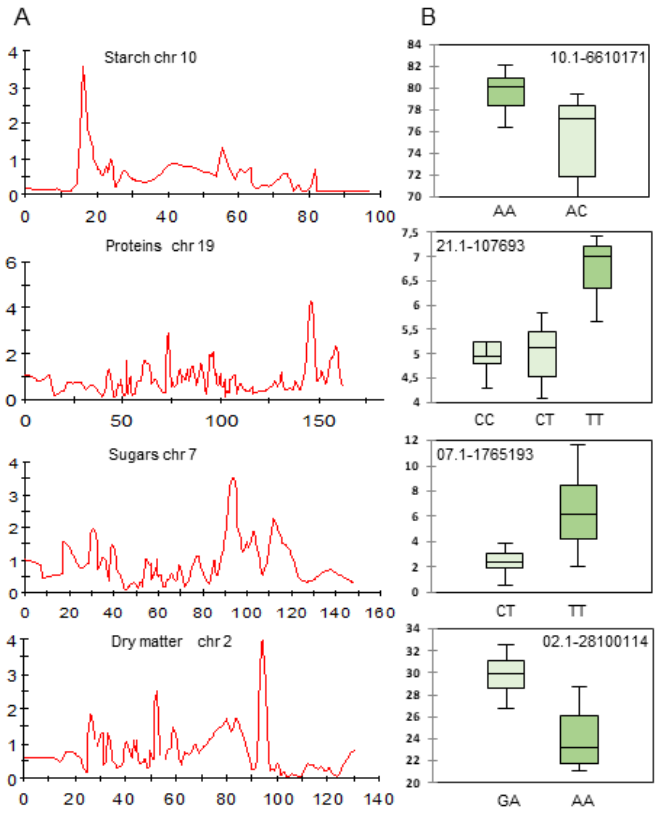


Figure 14: (A) Exemples de QTL détectés dans les populations biparentales, localisés sur les chromosomes 10, 19, 7 et 2 respectivement, pour l'amidon, les protéines, les sucres et la matière sèche. (B) Marqueurs SNP situés dans l'intervalle de confiance du QTL, présentant une association significative avec les données phénotypiques du panel de diversité. Les allèles à chaque locus et leurs effets alléliques sont représentés.

3. Validation des QTLs au sein d'un Panel de Diversité

Pour vérifier que les QTLs ne sont pas spécifiques au fond génétique des géniteurs parentaux, nous avons effectué une validation sur un panel de diversité de *Dioscorea alata* diversifié pour les divers traits étudiés. Le panel utilisé pour la validation des loci liés au traits morphologiques était composé de 28 accessions (tableau 10), tandis que celui qui a servi à la validation des loci physico-chimiques de 24 accessions. Ces accessions ont été soumises à un génotypage et à un phénotypage similaires à ceux des populations en ségrégation. La validation des QTLs, fondée sur les données génotypiques et phénotypiques des panels de diversité, s'est déroulée en deux étapes distinctes. Dans un premier temps, un modèle linéaire simple associant chaque phénotype à un SNP a été testé à l'aide du coefficient de corrélation de Pearson, avec un seuil de signification fixé à 5 %. Tous les SNPs inclus dans l'intervalle de confiance du QTL ont été testés. Dans un second temps, une analyse de variance (ANOVA) a été réalisée pour les

marqueurs significatifs afin de determiner si les classes génotypiques observées présentaient des différences significatives, avec un seuil de $p < 0,05$.

Tableau 10 : Détails des 28 accessions de *D. alata* du panel de diversité avec leur origine géographique, leur nom local, leur niveau de ploïdie, leur code d'accession et le type d'accession inclus dans l'étude.

Collection	Accession Code	Geographic origin	Local Name	Type of accession	Ploidy level
CRB-PT	PT-IG-00046	Puerto Rico	65_Sea 190	C	2
CRB-PT	PT-IG-00029	Haïti	47_Plimbite	C	2
CRB-PT	PT-IG-00350	Guadeloupe	403_Pacala	C	2
CRB-PT	PT-IG-00074	Barbados	93_Oriental	C	2
CRB-PT	PT-IG-00077	Puerto Rico	96_Kinabayo	C	2
CRB-PT	PT-IG-00565	Unknown	629_KL 21	C	2
CRB-PT	PT-IG-00373	Unknown	428_Hyb 5	C	2
CRB-PT	PT-IG-00019	New Caledonia	37_Gordito	C	2
CRB-PT	PT-IG-00023	Puerto Rico	41_Florido	C	2
CRB-PT	PT-IG-00049	Puerto Rico	68_Cinq	C	2
CRB-PT	PT-IG-00092	Puerto Rico	112_Caplaou	C	2
CRB-PT	PT-IG-00337	Guadeloupe	390_Boutou	BRL	2
CRB-PT	PT-IG-00335	Guadeloupe	388_A 9	BRL	2
CRB-PT	PT-IG-00349	Guadeloupe	402_A 22	BRL	2
CRB-PT	PT-IG-00361	Guadeloupe	416_A 17	BRL	2
CRB-PT	G15003	Guadeloupe	INRA15-Q_2b	BRL	2
CRB-PT	PT-IG-00076	New Caledonia	95_Bélep	C	3
CIRAD	Div-PB	Guadeloupe	Divin	C	2
CIRAD	VU423	Vanuatu	Manlankon	C	2
CIRAD	VU613	Vanuatu	Peter	C	2
CIRAD	VU528	Vanuatu	Sinoua	C	4
CIRAD	VU472	Vanuatu	Toufi Tetea	C	4
CIRAD	VU754	Vanuatu	Noulelcae	C	4
CIRAD	H4X431	Guadeloupe	Dou	BRL	4
CIRAD	H4X242	Guadeloupe	Roujol	BRL	4
CIRAD	H4X172	Guadeloupe	Hyb 4X-172	BRL	4
CIRAD	H4X274	Guadeloupe	Marchande	BRL	4
CIRAD	H4X131	Guadeloupe	Hyb 4X-131	BRL	4

C, Landraces; BRL, Breeding lines

En totalité, neuf QTL ont été validés sur le panel de diversité pour les traits morphologiques. Ces neuf QTLs, répertoriés dans le tableau 11, se décomposent en trois pour le ratio L/W, trois pour la régularité de la forme des tubercules, deux pour la texture de la peau, et un pour les racines tuberculaires. Les SNPs significatifs ainsi que l'effet des allèles sont également précisés. A titre d'exemple, la figure 15 offre une représentation visuelle de la distribution des valeurs phénotypiques en fonction des classes génotypiques observées dans le panel de diversité, pour trois de ces loci marqueurs.

Parmi les divers traits morpho-agronomiques influant sur la qualité du tubercule, le ratio longueur/largeur, la régularité de forme, et l'absence de racines tuberculaires revêtent une importance particulière. En effet, tant les producteurs que les consommateurs manifestent une préférence pour des tubercules légèrement allongés, exempts de digitations (formes

régulières), du fait de leur facilité de récolte et de stockage, ce qui réduit les risques de cassures et de blessures. De surcroît, ces tubercules sont plus faciles à éplucher lors de la préparation des repas. La présence de racines sur les tubercules, en plus de nuire à leur attrait visuel, complique également le processus d'épluchage. Enfin, la texture de l'épiderme des tubercules peut varier entre lisse et rugueuse. Les tubercules à la surface rugueuse présentent des défis supplémentaires lors de l'épluchage par rapport à ceux à la peau lisse, entraînant potentiellement des pertes significatives à cette étape.

Tableau 11: Caractéristiques des marqueurs SNP validés sur le panel de diversité.

	Chr	Position	SNP	Flanking sequences with SNP location
Ratio length/width				
16.1_23049864	16	23049864	T/A	ACCCATGGTTGGTGCAGAGGTGATTATGGGAACTGCATGG[T/A]TGGCTACGTTGGGTTTTCAATTTGATGAGTTATGAGTCACT
16.1_22716126	16	22716126	A/T	TACTAGACACGACATTGCTTATCTTTGTTAGCAATGTTAGC[A/T]GATTTCTCTCAAATCCAAGTAGAGAGCATTGGAATGCAG
06.1_16775930	15	16775930	A/T	TGCTAAGAGAGTTTTACCTAAGAAAGAAGAACTTACTCTC[A/T]GAGTTGCGAGGTTCCGTGTTTGATAGGCTTGTACTGCCGA
Shape regularity				
16.1_22654587	16	2654587	A/T	AGTGAATTTGTTGATCATGGTAGTGATGTTAAGTTTTTGT[A/T]ATAGCATAGCATTTGACTGTGTGTGGCCTGATTTATGCTC
16.1_22506101	16	22506101	T/A	ATGTACACACACACACACAAATAAAACACACAAATAGGTT[T/A]TAAAAATTTAAAGGAACGCTTGACATTGtAAAAATAATAA
02.1_19043957	2	19043957	C/G	TATATTGACATGATTCAATTCTACAATTTTATCTTCCTAT[C/G]TATTAATTAATAACTTCTATTTATAAAAAAATTGTAGTTT
Skin texture				
16.1_22564651	16	22564651	A/T	TAAAGACATAGACTTCACTGACTCTGTGGGGATAAGATAT[A/T]TGTTTTAGTTGGGTTTAATATTTGGTCCTATAAGAGTAAT
15.1_6502671	15	6502671	C/G	AGTAAACAGAAACCCAAACTATTAAACAGAAACTACTAAA[C/G]TTAATCAGATAAGCATTGACATATAATGTAAAACTATAAC
Tubercular roots				
16.1_21653648	16	21653648	C/T	GAGCTTAGTGCCCCAAAGAGGTAGTCGCCACTTCGCTCCT[C/T]TCCTTGGTGTACAACAGCACCATCGCTCCTCTCCATGTTAG

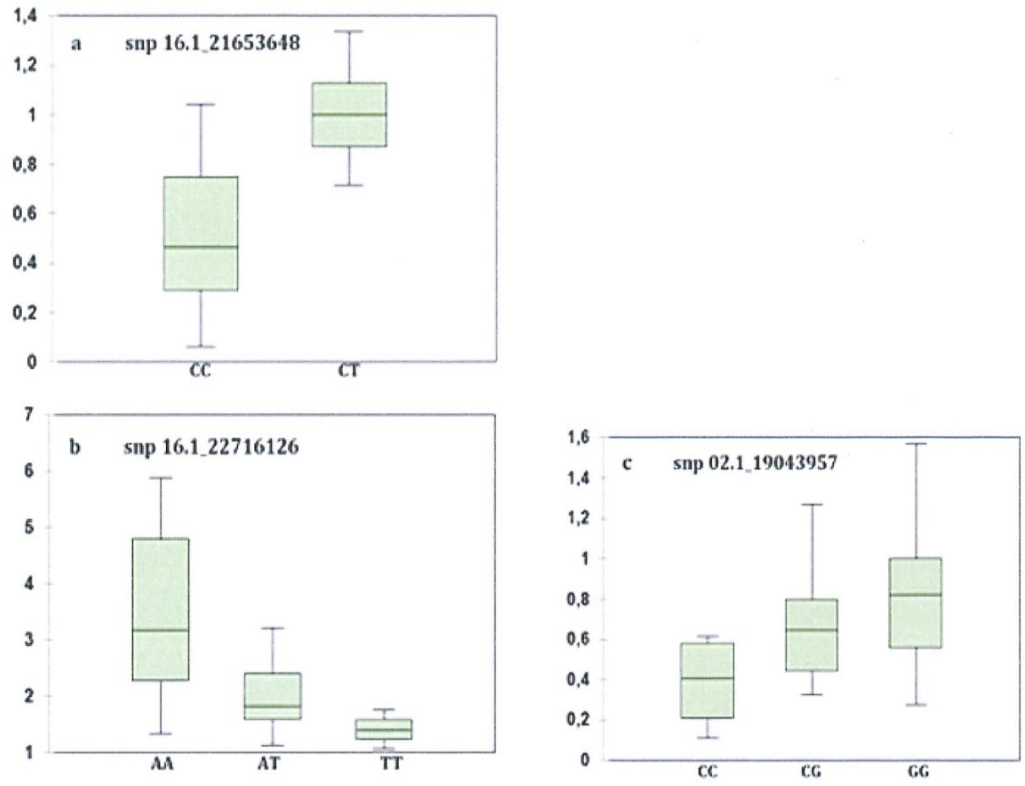


Figure 15 : Graphiques en boîtes à moustaches illustrant la distribution des données phénotypiques observées dans le panel de validation en fonction des classes génotypiques pour les locus 16.1_21653648, 16.1_22716126 et 02.1_19043957. 3a : Allèles régulant la quantité de racines tubéreuses, 3b : Allèles influençant le rapport longueur/largeur, 3c : Allèles influençant la régularité de la forme.

La validation des QTLs liées aux characteristiques physico-chimiques a été faite sur un panel de diversité composé de 24 accessions de *Dioscorea alata,* présentant des caractéristiques contrastées. Il s'agit des 24 accessions ayant servi au développent des modèles de prédiction NIRS, et dont les données phénotypiques avaient été obtenues par des analyses chimiques (Ehounou *et al*., 2021).

La majorité des QTL (22 sur 25) ont été validés sur le panel de diversité. Le tableau 9 présente les SNP situés dans les intervalles de confiance des QTL, montrant une association significative avec les données phénotypiques du panel de diversité (à p <0,01** ou p <0,05*). Les analyses de variance ont montré que les différences entre les différentes classes génotypiques étaient significatives (p<0,05) pour tous les QTL validés. La figure 14B illustre, à titre d'exemple, les distributions des données phénotypiques des différentes classes génotypiques pour les QTL détectés pour l'amidon, les protéines, les sucres et le DMC sur les chromosomes 10, 19, 7 et 2, respectivement, ainsi que les effets des allèles. Au locus 10.1.6610171, l'allèle C est associé à une teneur en amidon plus faible. Au locus 02.1.28100114, l'allèle A est associé à une faible teneur en matière sèche. Au locus 21.1.307693, trois classes génotypiques différentes ont été observées et les variétés homozygotes pour l'allèle T (TT) avaient des teneurs en protéines significativement plus élevées. Au locus 07.1.1765193, les variétés homozygotes pour l'allèle T (TT) avaient des teneurs en sucre significativement plus élevées.

En conclusion, les QTLs validés pourront être utilisés dans les programmes d'amélioration afin de sélectionner les allèles favorables pour ces caractères par le biais de la sélection assistée par marqueurs (SAM). Cette approche devrait se révéler prometteuse pour accélérer le processus de sélection des hybrides au sein des descendances en ségrégation.

4. Recherche de gènes candidats associés aux QTLs identifiés

Une recherche approfondie de gènes candidats a été menée dans les intervalles de confiance des divers QTLs ayant été validés, en utilisant le portail de génomique comparative du NCBI. Ce portail compile les informations relatives aux 35078 gènes annotés sur le génome de référence.

Pour la taille des tubercules, deux gènes candidats ont été identifiés. Le premier, situé près du SNP 16.1_23694121 (un gène du cycle de division cellulaire 202, cofacteur du complexe APC) et un second situé près des SNPs 16.1_22716126 et 16.1_22746342 (un gène régulateur du nombre de cellules 6). Le complexe APC est une ligase ubiquitine E3 et un régulateur essentiel de la progression du cycle cellulaire (Franks *et al.,* 2020). La famille des gènes CNR (régulateurs du nombre de cellules) est quant à elle connue pour influencer la taille des plantes et des organes (Guo *et al.,* 2010).

En ce qui concerne la présence de racines tuberculaires, un gène candidat a été repéré à proximité du SNP 16.1_21653548, lié à une protéine de liaison au promoteur de SQUAMOSA (SPL), qui impacte divers processus biologiques y compris le taux d'initiation des feuilles, la production de trichomes, la biosynthèse des anthocyanes et des caroténoïdes et le développement des racines latérales (Ma *et al*., 2022).

Pour la teneur en amidon, cinq gènes candidats potentiels ont été identifiés, dont quatre *(iso-amylase ISA3, sérine/thréonine protéine kinase, saccharose phosphate synthase, saccharose phosphatase*) jouent un rôle crucial dans le métabolisme de l'amidon et du saccharose. Il a été rapporté que l'Isoamylase ISA3 participe au processus de dégradation de l'amidon chez la pomme de terre (Ferreira *et al*., 2017). Cette enzyme est impliquée dans la réduction de la période de dormance (Ferreira *et al*., 2017). Le saccharose phosphate synthase et la saccharose phosphatase jouent des rôles clés dans le métabolisme du saccharose (Fernie *et al*. ,2002). La sérine/thréonine protéine kinase, a été signalé comme participant à la biosynthèse de l'amidon et du sucre chez la pomme de terre (Mckibbin *et al.,* 2006). Le cinquième gène putatif, la Glycosyl hydrolase de la famille 5, serait impliqué dans l'hydrolyse de la liaison glycosidique entre deux ou plusieurs glucides, et est présent dans de nombreux tissus végétaux (Minic, 2008 ;Neis et Da Silva Pinto, 2021).

Cinq gènes candidats potentiels ont été détectés pour la teneur en sucres, parmi lesquels trois jouent un rôle majeur dans la respiration ou la glycolyse de la plante (*Enolase, Pyruvate déshydrogénase kinase, Glycine H Protein*). La Pyruvate déshydrogénase kinase a été identifiée comme régulateur négatif de la pyruvate déshydrogénase mitochondriale, jouant un rôle central dans le contrôle de la respiration cellulaire (Millar *et al.,* 1998). L'Enolase, également connue sous le nom de Phosphopyruvate hydratase, catalyse la neuvième et avant-dernière étape de la glycolyse. Chez le tabac, il a été rapporté que la protéine H de glycine joue un rôle important dans le flux photorespiratoire (López-Calcagno *et al.,* 2019). La surexpression de cette protéine réduit la quantité de sucres solubles et augmenté l'accumulation d'amidon

(López-Calcagno *et al*., 2019). Le quatrième gène putatif,B*eta-glucosidase*, serait impliqué dans l'hydrolyse de la cellobiose et d'autres oligosaccharides en glucose (Angkasawati *et al.,* 2020). Le cinquième gène putatif, le transporteur UDP-rhamnose/UDP-galactose*,* serait impliqué dans le transport des sucres nucléotidiques du cytosol vers l'appareil de Golgi, pour être utilisés dans la synthèse de polysaccharides (Rautengartena *et al*., 2014).

En ce qui concerne la teneur en protéines, quatre gènes candidats potentiels ont été identifiés (*D-amino acid transaminase, Indole 3 glycerol phosphate synthase, Complex kinase 7, Proline rich like receptor kinase*), parmi lesquels la D-amino acid transaminase joue un rôle crucial dans le métabolisme de biosynthèse et/ou de dégradation de différents acides aminés chez les plantes, tels que l'alanine et la sérine (Diebold *et al.,* 2002). L'enzyme Indole 3 glycerol phosphate synthase, est impliquée dans la biosynthèse du tryptophane (Schlee et al., 2013). Tandis que l''enzyme Complexe Kinase 7 a été signalée comme étant impliquée dans la phosphorylation des protéines à la fois dans la membrane externe des mitochondries et dans les chloroplastes (Pagliarin et Dixon, 2006).

Enfin, trois gènes candidats potentiels ont été detectés pour la teneur en matière séche (*Xyloglucan galactosyltransferase, endo Beta D Glucanase, Ubiquitin protease*). Le premier est impliqué dans la plasticité des composants de la paroi cellulaire grâce à sa capacité à hydrolyser et à reconnecter les chaînes de xyloglucane (Baumann *et al.,*2007). Tandis que l'endo Beta D Glucanase, est impliquée dans le clivage des chaînes de glucanes, constituants majeurs des parois cellulaires générant principalement des oligosaccharides (Perrot et Ramirez 2022).

5.Projet de Recherche

Mon projet de recherche pour les prochaines années s'appuie sur les résultats obtenus jusqu'à présent et se concentrera sur deux axes principaux. D'une part, je prévois de valider fonctionnellement les gènes candidats identifiés par diverses analyses et méthodologies. D'autre part, je poursuivrai mes recherches sur la polyploïdie, visant à développer de nouvelles variétés capables de répondre aux défis agricoles actuels, notamment la baisse de la fertilité des sols et la diminution des rendements. Mes travaux antérieurs sur la polyploïdie m'ont convaincue de son intérêt potentiel pour le développement de variétés plus résilientes et plus productives.

Partie 1 : Caractérisation et validation de gènes candidats

1. Contexte et objectifs

Les avancées récentes en technologies de séquençage ont transformé les études génétiques, ouvrant la voie à des recherches approfondies et à une analyse plus précise des génomes végétaux. La mise au point de méthodes de séquençage de nouvelle génération (Next-Generation Sequencing, NGS) a permis de réduire drastiquement le coût du séquençage par rapport aux techniques de première génération, rendant ces analyses plus accessibles pour des projets de recherche (Satam *et al*., 2023). Grâce à ces technologies, le séquençage complet du génome (Whole Genome Sequencing, WGS) et la transcriptomique, autrefois onéreux et techniquement exigeants, sont désormais réalisables à grande échelle (Muñoz-Espinoza *et al* 2022 ; Song *et al*., 2023). La réduction des coûts permet aujourd'hui de générer des séquences pour de nombreux génotypes ou variétés, offrant une base de données riche et diversifiée pour identifier des variations génétiques spécifiques, telles que les SNPs et indels (insertions ou deletions), directement liées aux traits phénotypiques d'intérêt (Ramakrishna *et al.,* 2018; Liu *et al*., 2022).

Dans ce contexte, cette première partie de l'étude vise à valider les gènes candidats que j'ai identifiés, en confirmant leur rôle biologique et leur implication dans les variations phénotypiques des traits étudiés. Pour ce faire, j'utiliserai principalement des données de séquençage complet du génome (WGS) et de transcriptomique, développées par mon équipe DEFI.

2. Caractérisation des Variants par Séquençage Complet du Génome (WGS)

Dans un premier temps, j'utiliserai les données de séquençage WGS générées sur un large éventail de variétés (Mota *et al*., 2024), certaines ayant déjà été caractérisées pour les traits phénotypiques étudiés. Ces séquences génomiques seront alignées avec celles des gènes candidats identifiés pour explorer les variations génétiques, comme les SNPs et indels, en lien avec les traits observés. Les signatures spécifiques de certains allèles pourront servir de marqueurs génétiques dans les programmes de sélection variétale.

Ce type d'analyse nécessite des compétences en bio-informatique pour le traitement des données génomiques. Afin de répondre aux exigences de ces analyses complexes, j'ai suivi une formation spécialisée intitulée "Linux pour l'analyse des données génomiques" à l'Institut Agro de Montpellier (du 30 septembre au 4 octobre 2024, 35 heures). Cette formation m'a permis de maîtriser les outils bio-informatiques nécessaires pour le traitement et l'analyse des données de séquençage, en couvrant des aspects tels que la gestion des environnements Linux et l'utilisation de logiciels spécifiques pour chaque étape de l'analyse, y compris le mapping des reads sur les séquences des gènes candidats, l'appel et le filtrage des variants, ainsi que l'automatisation des flux de travail.

De plus, mon collègue B. Fouks a conçu un pipeline bio-informatique destiné à faciliter l'analyse des gènes candidats. Ce pipeline intègre de manière optimisée toutes les étapes nécessaires, depuis le traitement des séquences brutes jusqu'à l'identification et l'interprétation des variations génétiques (Figure 16), permettant ainsi d'optimiser les performances des analyses, leur reproductibilité, et la standardisation des résultats.

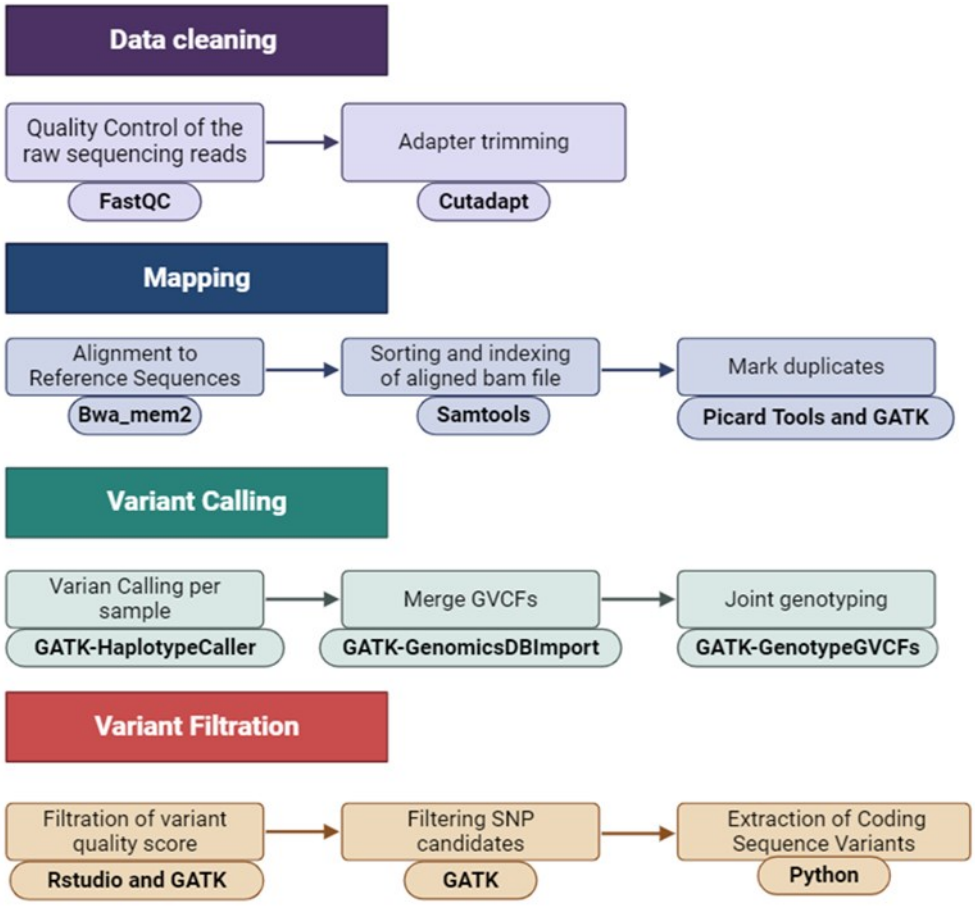


Figure 16 : Étapes et outils requis pour la caractérisation des gènes candidats à partir des données de séquençage complet du génome (Whole Genome Sequencing, WGS).

Si des variants sont détectés dans les régions codantes des gènes (CDS) et affectent les acides aminés, une validation fonctionnelle viendra confirmer leur impact biologique.

3. Validation par Étude Transcriptomique

Pour compléter l'analyse des variants génétiques, cette deuxième phase portera sur la validation fonctionnelle des gènes candidats via une étude transcriptomique. Les données, générées par mon collègue K. Dossa à partir d'échantillons de tubercules de six génotypes de Dioscorea alata (Mota *et al*., 2024), permettront une analyse comparative de l'expression génique. Les lectures transcriptomiques seront alignées avec les séquences des gènes candidats afin d'examiner si des différences d'expression sont en corrélation avec les variations phénotypiques. Ce travail visera à renforcer la compréhension du lien entre expression génique et caractéristiques observées.

4. Validation Fonctionnelle par qPCR

Afin de confirmer ces résultats, je réaliserai une validation par PCR quantitative (qPCR). Cette méthode mesurera l'expression des gènes d'intérêt dans des échantillons sélectionnés, afin de vérifier la régulation des différents allèles associés aux traits phénotypiques spécifiques. La qPCR contribuera à établir un lien solide entre les variants génétiques identifiés et les niveaux d'expression, renforçant les conclusions de l'étude sur le rôle de ces gènes dans les variations observées.

5. Validation par Silencing de Gène et Technologie CRISPR-Cas9

Enfin, si les résultats préliminaires sont concluants, je proposerai d'approfondir la validation des gènes candidats en utilisant des techniques avancées comme le silencing de gène (RNAi) ou CRISPR-Cas9. Ces méthodes permettront de moduler l'expression des gènes ciblés et d'en observer les effets sur les traits phénotypiques. Par exemple, le silencing de gène pourra être utilisé pour réduire ou supprimer l'expression de certains gènes, tandis que CRISPR-Cas9 introduira des mutations spécifiques pour en tester les impacts fonctionnels. Ces approches offriront une validation rigoureuse des rôles des gènes candidats, avec des applications prometteuses pour la sélection variétale.

Partie 2 : Augmentation du Rendement et de la Résistance aux Stress Abiotiques par la Polyploïdie

1. Contexte

La polyploïdie est un phénomène courant chez les plantes, qui a joué un rôle clé dans l'évolution et la diversification de nombreuses espèces cultivées (Soltis *et al*., 2004 ; Alix *et al.,* 2017 ; Tossi *et al*., 2022). Ce processus génétique permet souvent aux plantes polyploïdes de s'adapter à de nouveaux environnements (Dubcovsky et Dvorak, 2007 ; Tossi *et al*., 2022). Parmi les grandes cultures agricoles, beaucoup sont polyploïdes, notamment le blé (*Triticum aestivum*), l'avoine (*Avena sativa*), le coton (*Gossypium hirsutum*), le tabac (*Nicotiana tabacum*), le café (*Coffea arabica*), la pomme (*Malus pumila*), la poire (*Pyrus communis*), la patate douce (*Ipomoea batatas*), la luzerne (*Medicago sativa*) et la pomme de terre (*Solanum tuberosum*).

Parmi les espèces citées, les huit premières sont des allopolyploïdes, tandis que les deux dernières sont des autotétraploïdes. Les allopolyploïdes résultent de croisements interspécifiques suivis d'une duplication chromosomique, ce qui leur permet de cumuler plusieurs génomes. En revanche, les autopolyploïdes proviennent d'un seul génome ancestral qui a été dupliqué sans apport génétique d'autres espèces. Ce processus d'autopolyploïdie resulte souvent d'erreurs survenant lors de la méiose, menant à plusieurs ensembles de chromosomes homologues (par exemple, 3 copies = triploïdes, 4 copies = tétraploïdes).

La majorité des variétés cultivées de luzerne (*Medicago sativa*) (2n = 4x = 32) et de pomme de terre (*Solanum tuberosum*) (2n = 4x = 48) sont des autotétraploïdes, possédant ainsi quatre ensembles complets de chromosomes homologues. Ces caractéristiques polyploïdes apportent des avantages significatifs en matière de productivité agricole. Par exemple, des études ont montré que la luzerne tétraploïde présente une biomasse et un rendement supérieurs par rapport aux variétés diploïdes, ainsi qu'une meilleure capacité d'adaptation à des conditions environnementales variées (Dunbier *et al*., 1975; Bingham *et al*., 1994). Concernant la pomme de terre, il a été démontré que les variétés tétraploïdes augmentent le rendement global (Kumar et Kang, 2016). Cette amélioration des traits agronomiques est souvent attribuée à l'hétérosis, qui accroît la vigueur des plantes polyploïdes, les rendant plus résistantes aux stress abiotiques et biotiques (Mendoza et Haynes, 1974 ; Ortiz Ríos, 2015).

Ainsi, la polyploïdie ne constitue pas uniquement un phénomène évolutif, mais elle représente également un atout majeur pour l'agriculture moderne. Il a été démontré que, parmi les trois principales espèces d'igname, *Dioscorea rotundata*, *D. alata* et *D. trifida*, les deux dernières sont des autopolyploïdes. Trois cytotypes différents ont été identifiés (2n = 2x, 3x, 4x = 40, 60, 80) chez *D. alata* (Arnau *et al*., 2009) et *D. trifida* (Bousalem *et al*., 2006, 2010). Chez *D. trifida*, qui produit généralement plus d'un tubercule par plante, les formes tétraploïdes domestiquées génèrent des tubercules plus gros que les formes diploïdes (Bousalem *et al*., 2006). Ce phénomène est similaire à celui observé chez la pomme de terre, où la polyploïdie a conduit à une augmentation significative de la taille et du rendement des tubercules (Kumar et Kang, 2016).

Des travaux de polyploïdisation ont également été menés sur des espèces diploïdes telles que l'orge (*Hordeum vulgare*, 2n = 2x = 14) (Chen *et al*., 2021), le peuplier (*Populus hopeiensis*, 2n = 2x = 38) (Wu *et al*., 2023), le pommier (*Malus domestica*, 2n = 2x = 34) (Xue *et al.,* 2017) et le gingembre (*Zingiber officinale*, 2n = 2x = 36) (Zhou *et al.,* 2020). Ces études ont montré que les autotétraploïdes présentent plusieurs modifications phénotypiques par rapport aux diploïdes. Parmi ces modifications, on observe une augmentation de la taille des stomates (longueur et largeur), une diminution du nombre total de stomates, des feuilles plus grandes et une concentration accrue en chlorophylle (Chen *et al*., 2021; Wu *et al*., 2023; Xue *et al*., 2017; Zhou *et al.,* 2020). Ces différences phénotypiques s'expliquent par le doublement de l'ADN dans les tétraploides, entraînant une augmentation du volume cellulaire, favorisant des cellules plus grandes, dues à une accumulation de protéines et de métabolites essentiels. En outre, il a été démontré que la duplication du génome entraîne des modifications dans l'expression des gènes (Zhou *et al.,* 2015; Chen *et al*., 2021; Tossi *et al*., 2022). Dans les tétraploïdes, certaines voies métaboliques peuvent être amplifiées, ce qui affecte de manière significative la croissance et le développement phénotypique (Zhang *et al*., 2024).

En Guadeloupe, j'ai initié un programme de recherche sur la polyploïdie, mais plusieurs défis demeurent, notamment la rareté des tétraploïdes naturels et la difficulté à exploiter la production de gamètes diploïdes. Ce phénomène, observé chez certaines variétés mâles et femelles, est fortement influencé par les conditions environnementales. Pour remédier à cette situation, nous avons généré quelques tétraploïdes artificiels par duplication complète du génome (Whole Genome Duplication, WGD) à partir de variétés diploïdes sélectionnées comme géniteurs tétraploïdes.

Étant donné notre compréhension actuelle de l'architecture génétique des traits clés déterminant la qualité des tubercules, je propose d'élargir la recherche en vue de créer des hybrides tétraploïdes à haut rendement et résilients aux stress abiotiques. Pour cela, je prévois la production de tétraploïdes artificiels par duplication complète du génome à partir de 15 variétés diploïdes élite florifères, choisies pour la qualité de leur tubercule.

2. Objectifs

- Comparer les performances des diploïdes/diploides doublés : Évaluer l'impact de la polyploïdie sur des paramètres tels que le rendement, la taille et le nombre de stomates, ainsi que la composition biochimique des tubercules.
- Créer des hybrides tétraploïdes : Utiliser les tétraploïdes artificiels comme géniteurs pour produire des descendances maximisant l'hétérozygotie et l'hétérosis, en se basant sur la complémentarité des traits et la distance génétique entre les parents.

3. Sélection de Variétés Diploïdes Fertiles Supérieures

La première étape consistera à sélectionner des variétés diploïdes fertiles présentant des caractéristiques agronomiques intéressantes, notamment pour les traits qui influent sur la qualité des tubercules. Cette sélection s'appuiera sur les collections de germoplasmes, notamment celles du CRB-PT et du CIRAD. Nous utiliserons une combinaison de données

phénotypiques et de marqueurs moléculaires afin d'identifier des variétés diploïdes de haute qualité, qui serviront de base pour l'augmentation du niveau de ploïdie.

4. Production de Tétraploïdes à partir des Variétés d'Élite Diploïdes Sélectionnées

La deuxième étape consistera à induire le doublement chromosomique de quinze variétés diploïdes sélectionnées afin de les convertir en tétraploïdes, par la technique de duplication complète du génome (Whole Genome Duplication, WGD) à l'aide de traitements chimiques. Pour cela, j'ai établi un partenariat avec une société privée, reconnue pour son expertise dans l'induction de polyploïdie et les analyses de vérification par cytométrie en flux. Cette phase sera réalisée au cours de la première année.

5. Étude Comparative des Génotypes Diploïdes et Tétraploïdes

À partir de la deuxième année, nous évaluerons en Guadeloupe les performances au champ des 15 génotypes diploïdes et des 15 génotypes tétraploïdes, chacun testé sur 15 plants, soit un total de 450 plants. Cette étude analysera les rendements ainsi que divers traits agronomiques essentiels, tels que la teneur en chlorophylle des feuilles, la teneur en amidon des tubercules, la matière sèche et la teneur en protéines.

6. Croisements entre Tétraploïdes Complémentaires et Génétiquement Éloignés et Sélection des Meilleurs Hybrides

Dès la troisième ou quatrième année, des croisements pourront être effectués entre des tétraploïdes génétiquement éloignés et présentant des traits complémentaires. Ces croisements viseront à cumuler les allèles favorables associés à la qualité des tubercules tout en maximisant l'hétérosis pour améliorer le rendement. Les hybrides ainsi générés seront multipliés in vitro, testés en station, selon une approche combinant des évaluations phénotypiques et des marqueurs moléculaires, puis diffusés à nos partenaires pour des évaluations multilocales. Ce projet pourrait offrir des solutions innovantes pour améliorer la production de *Dioscorea alata* face aux défis environnementaux croissants et aux besoins alimentaires mondiaux.

Les activités envisagées sont déjà incluses dans deux projets de recherche en cours de montage, notamment un projet interrégional (projet CARIBIODIV) et le projet RTB Breeding IITA.

6. References bibiographiques

Abraham, K., & Nair, P. G. (1991). Polyploidy and sterility in relation to sex in *Dioscorea alata* L. (Dioscoreaceae). Genetica, 83, 93–97.

Abraham, K., & Arnau, G. (2007). Use of DNA markers (AFLP) for genetically improving the productivity, palatability, storability and dry matter content of tubers of greater yam. *Mid-term report of IFCPAR Project No. 3000 B 1*, period from 1 August 2005 to 27 May 2007. Retrieved from https://agritrop.cirad.fr

Abraham, K., & Arnau, G. (2010). Use of DNA markers (AFLP) for genetically improving the productivity, palatability, storability, and dry matter content of tubers of greater yam. Final-report of IFCPAR Project No. 3000 B 1. Retrieved from https://agritrop.cirad.fr

Abraham, K., Némorin, A., Lebot, V., & Arnau, G. (2013). Meiosis and sexual fertility of autotetraploid clones of greater yam *Dioscorea alata* L. *Genetic Resources and Crop Evolution*, 60(3), 819–823.

Alix, k, Gérard, P.R., Trude Schwarzacher, T., & Heslop-Harrison, J. S. (2017). Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants, Annals of Botany, 120, 2, 183–194, https://doi.org/10.1093/aob/mcx079

Angkasawati, V.S., Kamaliyah, I., Putra, MD., Nata, IF., & Irawan C .(2020). Structural characterization and antioxidant activity of liquid sugar from Alabio potato using enzymatic hydrolysis processes. IOP Conf Ser: Mater Sci Eng 980:012041 https://doi.org/10.1088/1757-899X/980/1/0120412.

Arnau, G., Maledon, E., Bachand, I., & Abraham, K. (2006). Production of interploid hybrids and molecular markers heterozygosity determination using microsatellite markers in the greater yam, *Dioscorea alata*: Importance for the genetic improvement of the greater yam. In *Proceedings of the 14th Triennial Symposium of the International Society for Tropical Root Crops* (pp. 20–26), Thiruvananthapuram, Kerala, India.

Arnau, G., Némorin, A., Maledon, E., & Abraham, K. (2009). Revision of ploidy status of *Dioscorea alata* L. (Dioscoreaceae) by cytogenetic and microsatellite segregation analysis. *Theoretical and Applied Genetics*, 118(7), 1239–1249.

Arnau, G., Abraham, K., Sheela, M. N., Chair, H., Sartie, A., & Asiedu, R. (2010). Yams. In J. E. Bradshaw (Ed.), *Root and tuber crops* (Vol. 7, Handbook of Plant Breeding, pp. 127–148). Springer.

Arnau, G., Bhattacharjee, R., Chair, H., Malapa, R., Lebot, V., Abraham, K., Perrier, X., Petro, D., Penet, L., & Pavis, C. (2017). Understanding the genetic diversity and population structure of yam (Dioscorea alata L.) using microsatellite markers. PloS One, 12(3), e0174150. 2.

Arnau, G., Desfontaines, L., Ehounou, A. E., Marie-Magdeleine, C., Kouakou, A. M., Leinster, J., Nudol, E., Maledon, E., & Chair, H. (2024). Quantitative trait loci and candidate genes for

physico-chemical traits related to tuber quality in greater yam (Dioscorea alata L.). Journal of the Science of Food and Agriculture, 104(8), 4872-4879.

Baumann, M.J., Eklöf, J.M., Miche,l G., Kallas, A.M., Teeri, T.T., Czjzek, M., et al., (2007) Structural evidence for the evolution of Xyloglucanase activity from xyloglucan Endo-Transglycosylases: biological implications for cell wall metabolism. Plant Cell 19:1947–1963 (). https://doi.org/10.1105/tpc.107.051391.

Bingham, E.T., Groose, R.W., Woodfield, D.R. and Kidwell, K.K. (1994), Complementary Gene Interactions in Alfalfa are Greater in Autotetraploids than Diploids. Crop Science, 34: 823-829.

Bousalem, M., Arnau, G., Hochu, I., Arnolin, R., Viader, V., Santoni, S., & David, J. (2006). Microsatellite segregation analysis and cytogenetic evidence for tetrasomic inheritance in the American yam *Dioscorea trifida* and a new basic chromosome number in the *Dioscoreae*. *Theoretical and Applied Genetics*, 113, 439–451.

Bousalem, M., Viader, V., Mariac, C., Gomez, RM., Hochu, I., Santoni, S, et al. (2010). Evidence of diploidy in the wild Amerindian yam, a putative progenitor of the endangered species Dioscorea trifida (Dioscoreaceae). Genome. 53(5):371–83

Bradshaw, .JE., Hackett, C.A., Pande, B., Waugh, R., & Bryan GJ. (2008). QTL mapping of yield, agronomic and quality traits in tetraploid potato (Solanum tuberosum subsp. tuberosum) Theor. Appl. Genet 116:193–211. doi: 10.1007/s00122-007-0659-1

Carputo, D., Barone, A., & Frusciante, L. (2000). 2n gametes in the potato: Essential ingredients for breeding and germplasm transfer. *Theoretical and Applied Genetics*, 101, 805–813.

Chen, Y., Xu, H., He, T., Gao, R., Guo, G., Lu, R., ... & Liu, C. (2021). Comparative analysis of morphology, photosynthetic physiology, and transcriptome between diploid and tetraploid barley derived from microspore culture. Frontiers in Plant Science, 12, 626916.

Cormier, F., Lawac, F., Maledon, E., Gravillon, M. C., Nudol, E., Mournet, P., Vignes, H., Chair, H., & Arnau, G. (2019). A reference high-density genetic map of greater yam (Dioscorea alata L.). Theoretical and Applied Genetics, 132(6), 1733-1744.

Diebold, R., Schuster, J., Daeschner, J.K & Binder, S. (2002). The branched-chain amino acid transaminase gene family in Arabidopsis encodes plastid and mitochondrial proteins. Plant Physiol 129:540–550.

Dubcovsky, J., & Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. Science, 316(5833), 1862-1866.

Dunbier, M.W., Eskew, D.L., Bingham, E.T. and Schrader, L.E. (1975), Performance of Genetically Comparable Diploid and Tetraploid Alfalfa: Agronomic and Physiological Parameters. Crop Science, 15: 211. ht

Ehounou, A. E., Cornet, D., Desfontaines, L., Marie-Magdeleine, C., Maledon, E., Nudol, E., Beurier, G., Rouan, L., Brat, P., Lechaudel, M., Nous, C., N'Guetta, A. S. P., Kouakou, A. M., & Arnau, G. (2021). Predicting quality, texture and chemical content of yam (*Dioscorea alata* L.) tubers using near infrared spectroscopy. Journal of Near Infrared Spectroscopy, 29(3), 128-139.

Ehounou, A. E., Cormier, F., Maledon, E., Nudol, E., Vignes, H., Gravillon, M. C., N'Guetta, A. S. P., Mournet, P., Chair, H., Kouakou, A. M., & Arnau, G. (2022). Identification and validation of QTLs for tuber quality related traits in greater yam (*Dioscorea alata* L.). Scientific Reports, 12(1), 8423.

Essad, S., & Maunoury, C. (1984). Variation géographique des nombres chromosomiques de base et polyploïdie dans le genre Dioscorea à propos du dénombrement des espèces transversa Brown, pilosiuscula Bert., et trifida L. Agronomie, 4(7), 611–617.

Fernie, A.R., Willmitze, L. & Trethewey, R.N. (2002). Sucrose to starch: a transition in molecular plant physiology. Trends Plant Sci 7:35–41 (https://doi.org/10.1016/s1360-1385(01)02183-5 PMID: 11804825.

Ferreira, S.J., Senning, M., Fischer-Stettler, M., Streb, S., Ast, M., Neuhaus, H.E. et al., (2017). Simultaneous silencing ofisoamylases ISA1, ISA2 and ISA3 by multi-target RNAi in potato tubers leads to decreased starch content and an early sprouting phenotype. PLoS One 12:e0181444 https://doi.org/10.1371/journal.pone.0181444.

Franks, J. L. *et al*., (2020). In silico APC/C substrate discovery reveals cell cycle-dependent degradation of UHRF1 and other chromatin regulators. PLoS Biol. 18(12), e3000975. https://doi.org/10.1371/journal.pbio.300097.

Guo, M. *et a*l., (2010). Cell Number Regulator1 affects plant and organ size in maize: implications for crop yield enhancement and heterosis. Plant Cell 22(4), 1057–1073. https://doi.org/10.1105/tpc.109.073676.

Hackett, C. A., & Luo, Z. W. (2003). TetraploidMap: Construction of a linkage map in autotetraploid species. *Journal of Heredity*, 94, 358–359.

Iwanaga, M., & Peloquin, S. J. (1982). Origin and evolution of cultivated tetraploid potatoes via 2n gametes. *Theoretical and Applied Genetics*, 61, 161–169.

Kumar, R., & Kang, G.S. (2016). Usefulness of Andigena (Solanum Tuberosum SSP. Andigena) Genotypes as Parents in Breeding Early Bulking Potato Cultivars. Euphytica 150, 107–115 https://doi.org/10.1007/s10681-006-9098-3

Liu, C., Tian, Y., Liu, Z.-x., Gu, Y.-z., Zhang, B., Li, Y.-h., Na, J. & Qiu, L.-j. (2022). Identification and characterization of long-InDels through whole genome resequencing to facilitate fine-mapping of a QTL for plant height in soybean (Glycine max L. Merr.). Journal of Integrative Agriculture, 21(7), 1903–1912.

López-Calcagno PE, Fisk S, Brown KL, Bull SE, South PF .,& Raines CA (2019). Overexpressing the H-protein of the glycine cleavage system increases biomass yield in glasshouse and field-grown transgenic tobacco plants. Plant Biotechnol J 17:141–151 https://doi.org/10.1111/pbi.12953 Epub 2018 Jul 22. PMID: 29851213; PMCID:PMC6330538.

Malapa, R., Arnau, G., Noyer, J. L., & Lebot, V. (2005). Genetic diversity of the greater yam (*Dioscorea alata* L.) and relatedness to *D. nummularia* Lam. and *D. transversa* Br. as revealed with AFLP markers. *Genetic Resources and Crop Evolution*, 52, 919–929.

Ma, L. *et al*., (2022). Characterization of squamosa-promoter binding protein-box family genes reveals the critical role of MsSPL20 in alfalfa flowering time regulation. Front. Plant Sci. 12, 775690. https://doi.org/10.3389/fpls.2021.775690

Marks, G. E. (1966). The enigma of triploid potatoes. *Euphytica*, 15, 285–290.

McKibbin, R.S., Muttucumaru, N., Pau,l M.J., Powers, S.J., et al., (2006). Production of high-starch, low-glucose potatoes through over-expression of themetabolic regulator SnRK1. Plant Biotechnol J 4:409–418.

Meijer, D., *et al*. (201). QTL mapping in diploid potato by using selfed progenies of the cross S. tuberosum×S. chacoense. Euphytica. 214:121. doi: 10.1007/s10681-018-2191-6.

Mendoza, H.A., & Haynes, F.L. (1974). Genetic basis of heterosis for yield in the autotetraploid potato. Theoret. Appl. Genetics 45, 21–25 https://doi.org/10.1007/BF00281169

Millar, AH., Knorpp, C.,& Leaver, C.J. (1998). Christopher JC and Hill SA, Plant mitochondrial pyruvate dehydrogenase complex: purification and identification of catalytic components in potato. Biochem J 334:571–576.

Minic, Z.(2008). Physiological roles of plant glycoside hydrolases. Planta 227:723–740.

Mota, A.P.Z., Dossa, K., Lechaudel, M. et al. (2024). Whole-genome sequencing and comparative genomics reveal candidate genes associated with quality traits in Dioscorea alata. BMC Genomics 25, 248 https://doi.org/10.1186/s12864-024-10135-2.

Muñoz-Espinoza, C., Di Genova, A., Sánchez, A., et al. (2020). Identification of SNPs and InDels associated with berry size in table grapes integrating genetic and transcriptomic approaches. BMC Plant Biology, 20, 365. https://doi.org/10.1186/s12870-020-02564-4

Neis, A., & da Silva Pinto L. (2021). Glycosyl hydrolases family 5, subfamily 5: relevance and structural insights for designing improved biomass degrading cocktails. Int J Biol Macromol 193:980–995 ISSN 0141-8130). https://doi.org/10.1016/j.ijbiomac.2021.10.062

Némorin, A., Abraham, K., David, J., & Arnau, G. (2012). Inheritance of tetraploid *Dioscorea alata* and evidence of double reduction using microsatellite marker segregation analysis. *Molecular Breeding*, 30(4), 1657–1667.

Némorin, A., David, J., Maledon, E., Nudol, E., Dalon, J., & Arnau, G. (2013). Microsatellite and flow cytometry analysis to help understand the origin of *Dioscorea alata* polyploids. *Annals of Botany*, 112(5), 811–819.

Ollitrault, P., & Jacquemond, C. (1994). Facultative apomixis, spontaneous polyploidization and inbreeding in Citrus volkameriana seedlings. Fruits, 49, 398–400.

Olson, M. S. (1997). Bayesian procedures for discriminating among hypotheses with discrete distributions: Inheritance in the tetraploid Astilbe biternata. Genetics, 147, 1933–1942.

Ortiz Ríos, R. (2015). Polyploidy and Plant Breeding. In: Plant Breeding in the Omics Era. Springer, Cham. doi.org/10.1007/978-3-319-20532-8_11.

Pagliarin, D.J & Dixon, J.E. (2006), Mitochondrial modulation: reversible phosphorylation takes center stage. Trends Biochem Sci 31:26–34 ISSN 0968–0004). https://doi.org/10.1016/j.tibs.2005.11.005.

Perrot, T, Pauly, M & Ramírez, V. (2022). Emerging roles of -Glucanases in plant development and adaptative responses. Plants 11:1119 https://doi.org/10.3390/plants11091119.

Ramakrishna, G., Kaur, P., Nigam, D., et al. (2018). Genome-wide identification and characterization of InDels and SNPs in Glycine max and Glycine soja for contrasting seed permeability traits. BMC Plant Biology, 18, 141. https://doi.org/10.1186/s12870-018-1341-2

Ramsey, J., & Schemske, D. W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics*, 29, 467–501.

Rautengartena, C., Ebert, B., Moreno, I., Templec, H., Hertera, T., Linke, B., et al. (2014). The Golgi UDP-Rhamnose / UDP-Galactose transporter family in Arabidopsis. Proc Natl Acad Sci U S A 111:11563–11568 https://doi.org/10.1073/pnas.1406073111.

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R.P., Banday, S., Mishra, A.K., Das, G., et al. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. Biology, 12, 997. https://doi.org/10.3390/biology12070997

Scarcelli, N., Dainou, O., Agbangla, C., Tostain, S., & Pham, J. L. (2005). Segregation patterns of isozyme loci and microsatellite markers show the diploidy of African yam *Dioscorea rotundata* (2n = 40). *Theoretical and Applied Genetics*, 111, 226–232.

Schlee, S., Dietrich, S., Kurćon, T., Delaney, P., Goodey, N.M., & Sterner, R. (2013). Kinetic mechanism of Indole-3-glycerol phosphate synthase. Biochemistry, 52, 132–142.

Soltis, D.E, Soltis, P.S., &Tate, J.A (2004). Advances in the study of polyploidy since Plant speciation. New Phytologist 161: 173–191.

Song, B., Ning, W., Wei, D., Jiang, M., Zhu, K., Wang, X., Edwards, D., Odeny, D.A., & Cheng, S. (2023). Plant genome resequencing and population genomics: Current status and future prospects. Molecular Plant, 16, 1252–1268. https://doi.org/10.1016/j.molp.2023.07.009

Tossi, V.E., Martínez Tosar, L.J., Laino, L.E., Iannicelli, J., Regalado, J.J., Escandón, A.S., Baroli, I., Causin, H.F., & Pitta-Álvarez, S.I. (2022). Impact of polyploidy on plant tolerance to abiotic and biotic stresses. Frontiers in Plant Science, 13. https://doi.org/10.3389/fpls.2022.869423

Van Ooijen, J.W. (2009). *MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species*. Kyazma BV, Wageningen, Netherlands.

Wu, J., Zhou, Q., Sang, Y., Zhao, Y., Kong, B., Li, L., & Zhang, P. (2023). In vitro induction of tetraploidy and its effects on phenotypic variations in Populus hopeiensis. BMC Plant Biology, 23(1), 557.

Xue, H., Zhang, B., Tian, J. R., Chen, M. M., Zhang, Y. Y., Zhang, Z. H., & Ma, Y. (2017). Comparison of the morphology, growth and development of diploid and autotetraploid 'Hanfu'apple trees. Scientia horticulturae, 225, 277-285.

Zhang, X., Zheng, Z., Wang, J. et al. (2024). In vitro induction of tetraploids and their phenotypic and transcriptome analysis in Glehnia littoralis. BMC Plant Biol 24, 439 https://doi.org/10.1186/s12870-024-05154-w.

Zhou, J., Guo, F., Fu, J. *et al*. (2020). In vitro polyploid induction using colchicine for Zingiber Officinale Roscoe cv. 'Fengtou' ginger. Plant Cell Tiss Organ Cult 142, 87–94. https://doi.org/10.1007/s11240-020-01842-1.

Zhou Y, Kang L, Liao S, Pan Q, Ge X, *et al*. (2015) Transcriptomic Analysis Reveals Differential Gene Expressions for Cell Growth and Functional Secondary Metabolites in Induced Autotetraploid of Chinese Woad (Isatis indigotica Fort.). PLOS ONE 10(3): e0116392. https://doi.org/10.1371/journal.pone.0116392.

7. Annexe : Tirés à part des principales publications

Publication 1

Nemorin, A., David, J., Maledon, E., Nudol, E., Dalon, J., & Arnau, G. (2013). Microsatellite and flow cytometry analysis to understand the origin of *Dioscorea alata* polyploids. *Annals of Botany*, 112(5), 811-819.

Annals of Botany **112**: 811–819, 2013
doi:10.1093/aob/mct145, available online at www.aob.oxfordjournals.org

ANNALS OF BOTANY
Founded 1887

Microsatellite and flow cytometry analysis to help understand the origin of *Dioscorea alata* polyploids

A. Nemorin[1], J. David[2], E. Maledon[1], E. Nudol[1], J. Dalon[1] and G. Arnau[1,*]

[1]*CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement), Station de Roujol, 97170 Petit Bourg, Guadeloupe, France and* [2]*UMR AGAP, Montpellier Supagro, 2, place Viala, 34060 Montpellier Cedex 2, France*
** For correspondence. E-mail gemma.arnau@cirad.fr*

Received: 26 February 2013 Returned for revision: 9 April 2013 Accepted: 15 May 2013 Published electronically: 1 August 2013

• *Background and Aims Dioscorea alata* is a polyploid species with a ploidy level ranging from diploid ($2n = 2x = 40$) to tetraploid ($2n = 4x = 80$). Ploidy increase is correlated with better agronomic performance. The lack of knowledge about the origin of *D. alata* spontaneous polyploids (triploids and tetraploids) limits the efficiency of polyploid breeding. The objective of the present study was to use flow cytometry and microsatellite markers to understand the origin of *D. alata* polyploids.
• *Methods* Different progeny generated by intracytotype crosses ($2x \times 2x$) and intercytotype crosses ($2x \times 4x$ and $3x \times 2x$) were analysed in order to understand endosperm incompatibility phenomena and gamete origins via the heterozygosity rate transmitted to progeny.
• *Results* This work shows that in a $2x \times 2x$ cross, triploids with viable seeds are obtained only via a phenomenon of diploid female non-gametic reduction. The study of the transmission of heterozygosity made it possible to exclude polyspermy and polyembryony as the mechanisms at the origin of triploids. The fact that no seedlings were obtained by a $3x \times 2x$ cross made it possible to confirm the sterility of triploid females. Flow cytometry analyses carried out on the endosperm of seeds resulting from $2x \times 4x$ crosses revealed endosperm incompatibility phenomena.
• *Conclusions* The major conclusion is that the polyploids of *D. alata* would have appeared through the formation of unreduced gametes. The triploid pool would have been built and diversified through the formation of $2n$ gametes in diploid females as the result of the non-viability of seeds resulting from the formation of $2n$ sperm and of the non-viability of intercytotype crosses. The tetraploids would have appeared through bilateral sexual polyploidization via the union of two unreduced gametes due to the sterility of triploids.

Key words: *Dioscorea alata*, endosperm balance, $2n$ gametes, bilateral sexual polyploidization, triploid origin, polyploidy.

INTRODUCTION

Dioscorea alata is a monocot that belongs to the family Dioscoreaceae. This gender includes >600 species (Ayensu and Coursey, 1972) of which the three main cultivated species are *D. rotundata*, *D. alata* and *D. trifida*. Yams are an important food crop in tropical and sub-tropical regions. They are dioecious herbaceous vines cultivated for their starchy tubers. They are exclusively propagated by vegetative multiplication by means of small tubers or small pieces of tubers. New combinations can be obtained via sexual reproduction, and breeding new cultivars has proven to be an efficient method for genetic improvement (Abraham and Nair, 1991; Egesi and Asiedu, 2002; Arnau *et al.*, 2010, 2011).

Dioscorea alata is a polyploid species with diploid ($2n = 40$), triploid ($2n = 60$) and tetraploid ($2n = 80$) cytotypes (Arnau *et al.*, 2009). The origin of this species is still a matter of debate because it has not yet been clearly identified in its wild state in nature, whereas wild forms have been mentioned by Shajeela *et al.* (2011). Ploidy increase is correlated with growth vigour, higher and more stable tuber yield and increased tolerance to abiotic and biotic stress (Malapa *et al.*, 2005; Lebot, 2009; Arnau *et al.*, 2010). Diversity studies have shown that the most common forms are diploids, followed by triploids, and that tetraploids are rare and only exist in diversification centres in Asia and the South Pacific (Abraham and Nair, 1991; Arnau *et al.*, 2009; Lebot, 2009). Recent studies have demonstrated a tetrasomic segregation for *D. alata* tetraploid clones ($2n = 4x = 80$) (Nemorin *et al.*, 2012). Autotetraploidy confers several advantages, of which hybrid vigour, also known as heterosis, is among the most common (Gallais, 2003; Parisod *et al.*, 2010). Autotetraploids can be formed from diploids in a variety of ways (Harlan and de Wet, 1975; Ramsey and Schemske, 1998), one of which is thought to involve the combination of two unreduced ($2n$) gametes (bilateral sexual polyploidisation, BSP) (Bretagnolle and Thompson, 1995; Ramsey and Schemske, 1998). There are two major ways to produce $2n$ gametes: by first division restitution (FDR) or by second division restitution (SDR) (Mok and Peloquin, 1975). These two cytological events have different genetic consequences. In fact, FDR $2n$ gametes transmit the major part of parental heterozygosity to progeny, whereas SDR $2n$ gametes are rather homozygous. Because BSP would only occur with the joint probability of two events of low likelihood, it is considered quite rare in natural populations (Husband, 2004). However, tetraploids have been obtained by BSP in several species such as *Dactylis glomerata* (Bretagnolle and Lumaret, 1995), *Trifolium pratense* (Parrott

and al., 1985) and diploid relatives of *Solanum tuberosum* (Mendiburu and Peloquin, 1977; Hutten *et al.*, 1995).

Alternatively, tetraploids may be produced in two steps via a triploid intermediary, through a process known as the triploid bridge (unilateral sexual polyploidization, USP). Tetraploids have been obtained by USP in several species such as *Cucumis sativus* (Diao *et al.*, 2009), *Achillea borealis* (Ramsey, 2007) and *Chamerion angustifolium* (Husband, 2004). Triploids that are formed through the union of a haploid (x) and a diploid unreduced ($2x$) gamete are of varying fertility depending on the species (Ortiz and Vuylsteke, 1995; Burton and Husband, 2001; Park *et al.*, 2002). Meiosis of triploids is irregular and results in a majority of aneuploid gametes. Euploid gametes that are haploid, diploid or triploid would be formed at the expected frequencies of $1/2^x$, $1/2^x + 1/2^x$, $1/2^x + 1/2^x + 1/2^x$, respectively (Lange and Wagenvoort, 1973), where x is the basic chromosome number. Depending on their fertility and the ploidy of their functional gametes (e.g. $n = x$, $2x$ or $3x$), triploids may produce tetraploids through the union of a balanced triploid gamete ($3x$) with a haploid gamete (x) of a diploid or a triploid. Non-gametic reduction in triploids can increase the frequency of balanced gametes ($2x$ and $3x$) by SDR and FDR, respectively, as is the case in *C. angustifolium* (Husband, 2004).

Unions of gametes of different ploidy levels are expected to interfere with seed development. In angiosperms, seeds are obtained by double fertilization (fertilization of the oosphere, producing the embryo, and fertilization of the central cell, leading to the formation of the endosperm). Because the central cell of most flowering plant species is homodiploid ($2x$) and fertilized by a haploid male gamete (x), the resulting endosperm is triploid ($2 + 1$) and, therefore, genetically distinct from the diploid embryo ($1 + 1$). The endosperm of most angiosperms is triploid, with a 2:1 ratio of the maternal to the paternal genome, although exceptions have been found in some species (Williams and Friedman, 2002; Mizuochi *et al.*, 2009). Endosperm is particularly sensitive to ploidy unbalance (Köhler *et al.*, 2010). An unbalanced endosperm may lead to an unviable seed (Costa *et al.* 2004). Deviations from the ratio of two maternal (2m) to one paternal (1p) genome in the endosperm can cause endosperm failure (Köhler *et al.*, 2010). Increased contributions of maternal or paternal genomes inhibit proliferation of the endosperm or cause endosperm excess, respectively (Scott *et al.*, 1998).

Dioscorea alata breeding programmes were exclusively based on the creation of diploid varieties until 2006, although polyploidy has been acknowledged for a long time (Ramachandran, 1968; Abraham and Nair, 1991). The first polyploid hybrids were recently created by conventional hybridization, thanks to the discovery of the fertility of *D. alata* tetraploid varieties and to the development of an *in vitro* immature embryo rescue method (Arnau *et al.*, 2006; Abraham and Arnau, 2009). Intercytotype crosses between diploid females and tetraploid males revealed the phenomenon of endosperm incompatibility because they produce non-viable mature seeds with shrivelled endosperm. Obtaining tetraploid hybrids from intercytotype crosses suggested the ability of diploid progenitors to produce unreduced ($2n$) gametes, pollen or ovules. In the case of *D. alata*, triploid males are sterile because their flower buds remain unopened until drying off (Abraham and Nair, 1991), and the fertility of female flowers has never been demonstrated.

Flow cytometry is a fast and easy technique to determine ploidy levels in plants (Doležel *et al.*, 1994; Seker *et al.*, 2003) and has already been used to screen $2x \times 2x$ *D. alata* progenies (Arnau *et al.*, 2011). Flow cytometry has also been used to determine endosperm ploidy in several species and to better understand the phenomenon of endosperm incompatibility in interspecific or intercytotype crosses (Pichot and Maataoui, 1997; Sliwinska *et al.*, 2005).

Because microsatellite markers are co-dominant and highly reproducible, they are suitable for the analysis of allele segregation in progenies (Ashley and Dow, 1994). Once maternal and paternal microsatellite genotypes are known, the different pathways for gamete production and unions that give rise to polyploid individuals can be compared for their likelihood. For example, the origin of $2n$ gametes (maternal or paternal) can be deduced from microsatellite analysis. The transmitted heterozygosity can also provide knowledge about the type of mechanisms involved in the non-reduction, i.e. the suppression of restitution of the first or the second division.

Implication of gametic non-reduction in the formation of polyploid individuals has never been demonstrated in *D. alata*. Furthermore, endosperm incompatibilities in this species have never been studied, in spite of their importance in intercytotype crosses. This knowledge is crucial for the optimization of the production of new polyploid ($3x$ and $4x$) cultivars.

The objective of the present study was to use flow cytometry and microsatellite markers to understand the origin of *D. alata* spontaneous triploids and tetraploids. Diploid genitors suspected of producing unreduced gametes were used. Intercytotype crosses between diploid, triploid and tetraploid genitors generated different types of progeny. Flow cytometry was used to measure embryo and endosperm ploidy. Microsatellite markers were genotyped on diploid parents and their detected polyploid offspring. This work made it possible to identify the origin of $2n$ gametes and endosperm, and provided knowledge about the formation of *D. alata* polyploids.

MATERIALS AND METHODS

Plant materials

The origin of the *Dioscorea alata* plant material used is given in Table 1.

The first progeny of 300 seeds was obtained by crossing two diploid parents ($2n = 2x = 40$) (female '5F' and male 'Kabusa'). The diploid status of '5F' and 'Kabusa' was ascertained by flow cytometry in Arnau *et al.* (2009). Both diploid genitors are suspected of producing unreduced gametes. Half of the seeds ($n = 150$) were sown, and fresh leaves of seedlings were analysed using flow cytometry. The other half of the seeds were desiccated 90 d after pollination and the endosperm

TABLE 1. *Origin of plant material*

Plant	Type	Origin	Ploidy
'5F'	Hybrid	Guadeloupe	$2x$
'Kabusa'	Local cultivars	Caraïbes	$2x$
258F	Local cultivars	Madagascar	$3x$
148	Local cultivars	Vanuatu	$4x$

was separated from the embryo. Flow cytometry was then performed on seedlings resulting from embryo culture, whereas a joint flow cytometry analysis was carried out on endosperm. Simple sequence repeat (SSR) analyses were only performed on the leaves of detected triploid seedlings.

The second progeny of 2000 seeds obtained by crossing the diploid female clone '5F' and the tetraploid male clone 148 was used: (1) to study endosperm incompatibility in $2x \times 4x$ crosses; and (2) to analyse events of non-gametic reduction in the female parent. One thousand seeds were sown to evaluate seed viability and to check the ploidy of emerging plantlets by flow cytometry. The remaining 1000 seeds were desiccated to allow a separate ploidy analysis on endosperm and embryos.

The gametic fertility of a triploid female clone (258F) was analysed by crossing it with the diploid male 'Kabusa'. The fertility study of the male 'Kabusa' was done by staining the pollen with carmine. Obtaining tetraploids by this cross would be indicative of the formation of unreduced balanced gametes in triploid females. A total of 300 female flowers were fertilized by manual hybridizations, corresponding to 1800 potential seeds, given that a fruit can contain 1–6 seeds. The rates of fruit set and seed setting were recorded.

Flow cytometry analysis

Flow cytometry analysis of leaves was performed as described by Arnau *et al.* (2009). The nuclear DNA content of samples was determined by comparison of the relative positions of the G_{0-1} peak of different internal references: 760a as the triploid reference, 639a as the diploid reference and 754a as the tetraploid reference. The ploidy of these three clones was determined by mitotic chromosome counts in Arnau *et al.* (2009). For endosperm analysis, the protocol described by Sliwinska *et al.* (2005) was used, with some adaptations. Endosperm was chopped up with a double-edged razor blade in 1 mL of nucleus isolation buffer [0·1 M Tris–HCl, 2·5 mM $MgCl_2 \cdot 6H_2O$, 85 mM NaCl, 1 % (w/v) polyvinylpyrrolidone-PVP-10 and 0·1 % (v/v) Triton X-100 pH 7]. The suspension was filtered through a 30 μm pore filter. A 300 μL aliquot of filtrated endosperm solution, 200 μL of filtrated leaf solution of an internal standard and 400 μL of isolation buffer supplemented with propidium iodide (50 μg mL^{-1}) and RNase A (50 μg mL^{-1}) were then mixed in a tube. The suspensions were incubated for approx. 5 min at room temperature. After incubation, each sample was run on a flow cytometer. DNA quantities were measured using a FACScalibur laser flow cytometer (Beckton Dickinson, USA) with Cellquest Software.

The choice of the internal reference was made according to the expected offspring or endosperm ploidy. To detect non-expected seedlings (triploid in a diploid population or tetraploid in a triploid population), an appropriate internal reference [diploid (639a) or triploid (760a), respectively] was used and the results were interpreted as follows. The internal reference produces two fluorescence peaks: a major peak corresponding to $2x$ or $3x$ DNA quantities of the majority of leaf cells, and a replicated G_2 minor peak corresponding to $4x$ or $6x$ DNA quantities from cells in mitotic interphase. When the fluorescence peak corresponding to the nuclei obtained from a given sampled individual was between these two reference peaks, the individual was assumed to be triploid with the diploid internal reference or tetraploid with the triploid internal reference if the sample peak is closer than the G_{0-1} peak. When no additional peak was observed between the two peaks of the internal reference, the individual was assumed to have the same ploidy level as the reference. The same principle is applied to measure expected hexaploid ($6x$) endosperm. The tetraploid (754a) internal reference was used and produced two fluorescence peaks. When the fluorescence peak corresponding to the nuclei obtained from a given sampled endosperm was equidistant from these two reference peaks, the endosperm was assumed to be $6x$.

Microsatellite amplification

Seventy-five SSRs developed from five different yam species – *D. rotundata*, *D. abyssinica*, *D. prahensilis*, *D. japonica* and *D. alata* (Misuki *et al.*, 2005; Tostain *et al.*, 2006; Andris *et al.*, 2010) – were used to determine the genotypes of the genitors. Only the six SSR markers that revealed polymorphism between the diploid parents '5F' and 'Kabusa', without any common allele (Da2F10, Da1D08, mDaCIR8, mDaCIR60, mDaCIR61 and mDrCIR128), were selected. Primer sequences are given in Table 2. Da2F10, mDaCIR8, mDaCIR61 and mDrCIR128 show two alleles in the female progenitor '5F' and two different alleles in the male progenitor 'Kabusa'. Da1D08 and mDaCIR60 show heterozygosity in '5F' and are homozygous in 'Kabusa' but with different alleles.

Amplification was performed in a total volume of 20 μL containing 0·05 U μL^{-1} of *Taq* polymerase, 2 μL of 10× buffer, 0·2 mM dNTP, 2 μmol of labelled or unlabelled primers, 2 mM $MgCl_2$ and 10 ng of DNA. Forward primers were labelled with one of the following fluorophores: TET, NED, HEX or 6-FAM. The PCR conditions were as follows: 5 min of denaturation at 94 °C, followed by 30 cycles alternating 30 s of denaturation at 94 °C, 30 s of hybridization at annealing

TABLE 2. *Primer sequence of the six microsatellite markers using for study of the allelic contribution of the progenitor*

Locus	Sequence F	Sequence R	Origin
Da2F10	TCAAGGATAAGAACTCCC	CAACGGCTAAACAGAAA	*D. alata*
Da1D08	GATGCTATGAACACAACTAAC	TTTGACAGTGAGAATGGA	*D. alata*
mDaCIR8	ACAGCAGCAAAATAACTG	TCTTTGCAGGAGAAGAGG	*D. alata*
mDaCIR60	CAAAGACCAGGGAATGTG	AGAATGCAGAGCATGGTG	*D. alata*
mDaCIR61	CTAACCCCTCCAAAGCTG	GGGCATTCACGTCTTTAT	*D. alata*
mDrCIR128	CCGTATTCCAAGCGATAA	AGCGTGAAAACCTGATAAAA	*D. rotundata*

The forward sequence (sequence F) and reverse sequence (sequence R) are given.

temperature, 35 s of extension at 72 °C, and ending with 5 min of final elongation at 72 °C. PCR was carried out using a PTC100 thermocycler (MJ Research). Electrophoregrams were obtained by migration of amplification products on an ABI PRISM-TM 3100 automatic sequencer (Applied Biosystems). Allelic profiles were determined using GeneMapper v3·7 software (Applied Biosystems). Parents and polyploid zygotes eventually found in 150 seeds (second seed lot) of the '5F' × 'Kabusa' population were genotyped.

RESULTS

Triploids from the 2x × 2x *cross: '5F' × 'Kabusa'*

Frequency. Out of the 300 plantlets or embryos of the '5F' × 'Kabusa' progeny, four were found to be 3*x* (1·3 %) and 296 were found to be 2*x* (98·7 %). The diploid internal reference 639a was used for flow cytometry.

The aspect of all the sown seeds that produce seedlings (diploid and triploid) was normal (plump) (Fig. 1A). No clear morphological differences made it possible to differentiate *a priori* 3*x* seeds from 2*x* seeds.

Endosperm ploidy and gamete origin. Out of the 150 seed lot for which endosperm was detached, flow cytometry was possible on two seeds that then produced 3*x* plants, hereafter referred to as 5K1 and 5K5. The diploid internal reference 639a was used for flow cytometry. Endosperm peaks for the two seeds corresponded to a 3*x* state (Fig. 2). Two other seeds that also produced 3*x* embryos could not be analysed because of a shrivelled endosperm (Fig. 1B). These two 3*x* individuals were obtained by embryo rescue and are hereafter referred to as 5K11 and 5K34.

Parental ('5F' and 'Kabusa'), 5K1, 5K5, 5K11 and 5K34 genotypes for the six microsatellite markers are given in Table 3.

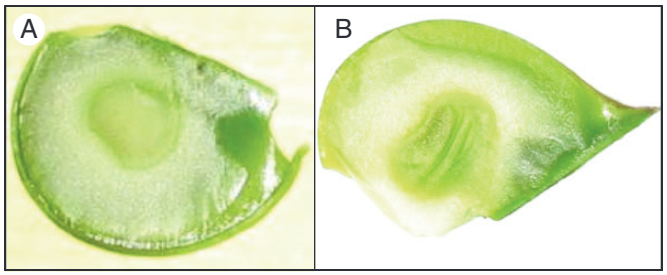


FIG. 1. Photographs of a plump seed in (A) and a shrivelled seed in (B).

Microsatellite analysis conclusions on the origin of the 3*x* individuals were drawn as follows. When the 3*x* offspring carried the two alleles of the female ('5F') and only one allele from the male ('Kabusa'), it was assumed that the female produced an unreduced ovule and the male a normal reduced pollen grain. When the two paternal alleles and only one maternal allele were transmitted, the unreduced gamete was assumed to be a pollen grain. In Fig. 3, the case of the 5K34 seedling is illustrated for locus mDaCIR8. The two male alleles (166 and 184 bp) are present in 5K34 with only one of the maternal alleles (171 bp). It can be postulated at this locus that 5K34 resulted from an unreduced pollen grain and a normal ovule. For other loci (mDaCIR61 and mDrCIR128), heterozygosity was not transmitted by the male or by the female.

Identically, triploid individuals obtained from plump seeds, 5K5 and 5K1, received heterozygosity from '5F' on loci Da2F10 and Da1D08. It can be hypothesized that they arise from a non-reduced ovule (Table 3). Triploids obtained from shrivelled seeds, 5K34 and 5K11, are assumed to come from non-reduced pollen grains.

They are not the same loci that transmit male or female heterozygosity. On average, the rate of transmitted heterozygosity is 50 % for 2*n* ovules (the case of 5K1 and 5K5) and approx. 22 % for 2*n* pollen (5K11 and 5K34).

Offspring of the female 2x × *male* 4x *cross: '5F' × 148*

Out of the 1000 sown seeds obtained by crossing the diploid female ('5F') and the tetraploid male (148), 18 seedlings

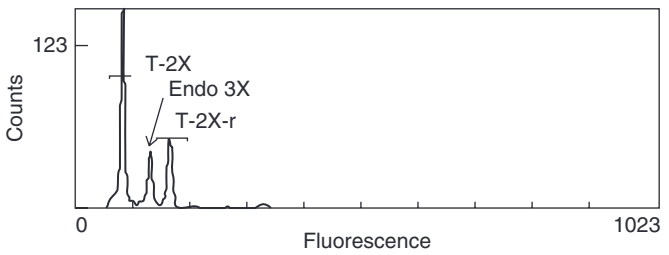


FIG. 2. Flow cytometry analysis of detached endosperm from desiccated seeds from which triploid plants grew. Seeds were obtained on the '5F' × 'Kabusa' cross, both 2*x* parents. The diploid internal reference (639a) produced two fluorescence peaks: T-2X, the major G_{0-1} peak corresponding to a 2*x* DNA quantity of the majority of leaf cells; and T-2X-r, the replicated G_2 minor peak corresponding to the 4*x* DNA quantities of cells in mitotic interphase. The fluorescent peak corresponding to endosperm (Endo 3X) is equidistant to the two reference peaks and is triploid (3*x*).

TABLE 3. *Parental and hybrid genotypes at six SSR markers*

	Da2F10	Da1D08	mDaCIR8	mDaCIR60	mDaCIR61	mDrCIR128
'5F'	126/132	303/313	171/188	146/157	200/217	286/310
'Kabusa'	124/130	306	166/184	142	188/193	300/308
5K5	_	_	166/171	142/146/157	188/200/217	308/310
5K34	_	303/306	166/171/184	142/146	188/200	300/310
5K1	124/126/132	303/306/313	166/171	_	_	300/310
5K11	124/132	_	166/184/188	142/146	_	286/308

The female progenitor '5F', the male progenitor 'Kabusa' and four of their offspring, 5K1, 5K5, 5K11 and 5K34, were analysed (out of a 300 progeny). Alleles are given in bp.

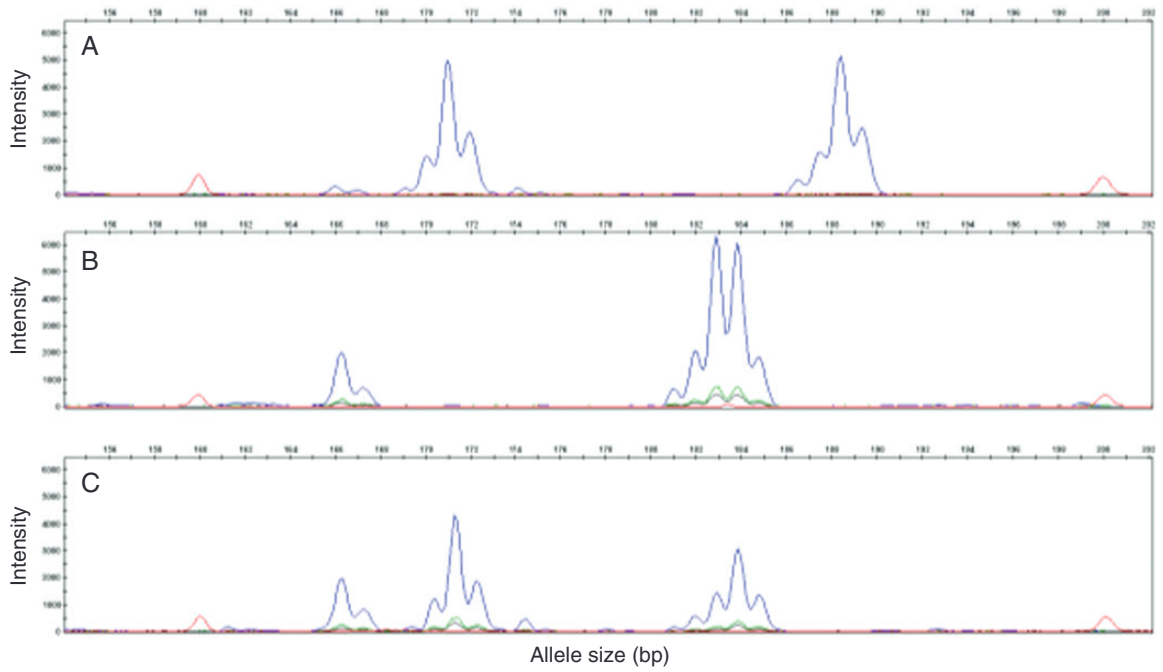


FIG. 3. Phenomenon of non-reduction of the male gamete. Electrophoregram at locus CIR8 (A) for the diploid female progenitor '5F', (B) for the diploid male progenitor 'Kabusa' and (C) for one triploid hybrid 5K34 with shrivelled endosperm. '5F' have phenotype 171/188; 'Kabusa' have phenotype 166/184; 5K34 have phenotype **166**/171/**184** (bold indicates the paternal allelic contribution).

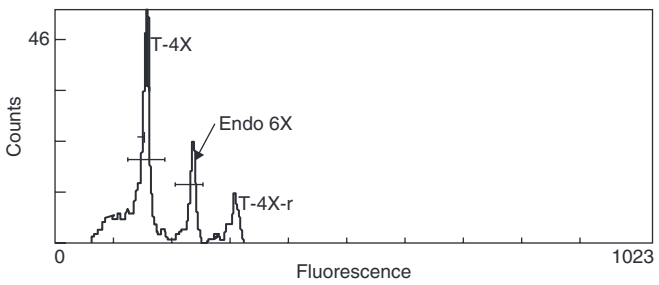


FIG. 4. Flow cytometry analysis of detached endosperm from desiccated seeds from which tetraploid plants grew. Seeds were obtained on the '5F' ($2x$) × 148 ($4x$) cross. The tetraploid internal reference (754a) produced two fluorescence peaks: T-4X, the major G_{0-1} peak corresponding to a $4x$ DNA quantity of the majority of leaf cells; and T-4X-r, the replicated G_2 minor peak corresponding to the $8x$ DNA quantities of cells in mitotic interphase. The fluorescent peak corresponding to endosperm (Endo 6X) is equidistant to the two reference peaks and is hexaploid ($6x$).

germinated from seeds with a plump aspect. Flow cytometry showed that these seedlings were all tetraploids. The triploid internal reference was used.

Out of the 1000 dessicated seeds, 995 had a shrivelled aspect and five were plump. Flow cytometry was carried out on 100 of the 995 shrivelled detached endosperms, but only six had a sufficient signal. These six endosperms had a signal peak between the two reference peaks of the triploid internal reference '760a' ($3x$ and $6x$ for the G_2 peak). They were therefore assumed to be $4x$. The corresponding embryos of these 100 shrivelled seeds were positioned as triploids using flow cytometry. Flow cytometry analysis showed that embryos of the five plump seeds were tetraploid. The endosperm peak of these five seeds was located at a $6x$ position (Fig. 4) when the tetraploid internal reference was used.

Fertility of triploid females: 258F × 'Kabusa'

Results of fertility study on the male 'Kabusa' showed that the pollen viability is high (80 %). A very small number of seeds was obtained by crossing the triploid female 258F with the diploid male 'Kabusa' (only 18 out of an optimal number of 1800 seeds). Half (nine) of these seeds lacked an embryo. The other half contained embryos but none of them evolved into a viable seedling, even if development was initiated in two of them.

DISCUSSION

This work is the first study reported in Dioscoreacea where seeds were desiccated to allow separate ploidy analysis on endosperm and embryos. However, it appeared that although endosperm analysis is efficient for plump seeds, it is not easy to apply to shrivelled seeds. Because polyploids were obtained at a low rate, this could lead to doubts about the sample size of the progeny used to obtain accurate estimates of non-reduced gamete formation. Our results should be considered more as qualitative rather than accurate estimates.

In this study, $2x \times 2x$ crosses gave 1·3 % of $3x$ individuals with a contrasted endosperm aspect. The flow cytometry of joint embryos and endosperms, seed aspects and microsatellite analysis together led to evidence that both male and female clones produced $2n$ gametes that mated with x gametes to produce viable $3x$ plants. When the unreduced gamete was an ovule, endosperm ploidy was $3x$ and the seed was plump.

When the unreduced gamete was a pollen grain, endosperm was shrivelled and its ploidy level could not be determined.

In the same manner, 4*x* individuals were obtained in the 2*x* × 4*x* cross (1·15 %). These seeds are easily identified by the plump aspect of their seeds, corresponding to the expected 6*x* endosperm if the diploid female produced a 2*n* ovule and if the endosperm originated from the fusion of the two 2*n* cells of the embryo sac (2*x* + 2*x*) with the normal 2*x* male gamete produced by the 4*x* male. For the other seeds, embryos were 3*x* and the endosperm was shrivelled, consistent with the union of a normal reduced ovule (*x*) from the diploid female and a reduced pollen grain from the 4*x* male.

This suggests that the production of 2*n* gametes is at the origin of *D. alata* polyploids, as is the case for potatoes (Iwanaga and Peloquin, 1982; Carputo *et al.*, 2000) and many other species (for a review, see Ramsey and Schemske, 1998).

Other phenomena such as polyspermy (fertilization of an egg by two sperm nuclei) and endospermal polyembryony (formation of embryos from endosperm cells) have been suggested as possible mechanisms that could lead to the production of triploids in other species. Polyspermy is not considered as a likely mechanism in the formation of polyploids (Harlan and DeWet, 1998) but has been demonstrated in the genus *Juglans* (Navashin and Finn, 1951). In *D. alata*, neither polyspermy nor polyembryony can explain the microsatellite profiles obtained on the triploids for the six loci tested. Under the polyspermy hypothesis, regardless of the paternal genotype, individuals would carry only one of the paternal alleles at all loci because the two sperm nuclei are identical by mitosis (Borg *et al.*, 2009). For some loci, paternal heterozygosity in triploids was observed and makes it possible to eliminate polyspermy. Endospermal polyembryony, which consists of the formation of embryos from endosperm cells, was observed in the genera *Bracharia* (Muniyamma, 1977), *Beta* (Yarmolyuk *et al.*, 1990) and *Citrus* (Gmitter *et al.*, 1990). These examples are rare and weakly substantiated (Batygina and Vinogradova, 2007). In this case, regardless of the maternal genotype, individuals would carry only one maternal allele at all loci since the two polar nuclei are identical by mitosis. Because maternal heterozygosity was observed in our *D. alata* triploid individuals at some loci, polyembryony may also be eliminated.

All these elements make it possible to conclude that triploid *D. alata* originate from 2*n* gamete formation. Two basic types of meiotic restitution mechanisms that lead to 2*n* gamete formation have been reported: FDR and SDR (Mok and Peloquin, 1975; Park *et al.*, 2007). These two cytological events do not transmit the same parental heterozygosity to their progeny. FDR mechanisms lead to 2*n* gametes that contain non-sister chromatids between the centromere and the first crossover. Consequently, all loci between the centromere and the first crossover that were heterozygous in the diploid parent will be heterozygous in the 2*n* gametes. Half of those beyond the crossover will be heterozygous in the gametes. SDR mechanisms lead to 2*n* gametes that contain sister chromatids between the centromere and the first crossover. All loci between the centromere and the first crossover that were heterozygous in the diploid parent will be homozygous in these 2*n* gametes, whereas those beyond the crossover will be heterozygous (Peloquin *et al.*, 2008). As a result, heterozygosity transmission by FDR varies from 100 to 50 %, and transmission by SDR varies from 0 to 100 % and depends on the position of markers in relation to the centromere (Park *et al.*, 2007).

Unreduced gametes were observed in the two sexes with a similar low probability (<2 %). Microsatellite markers were used to detect non-gametic reduction in males and females by studying heterozygosity transmission. Our preliminary results would indicate that unreduced ovules could transmit more heterozygosity than unreduced pollen. However, our study clearly lacks the statistical power to draw a final conclusion: a greater number of 3*x* individuals should be genotyped on a greater number of SSRs. Further analysis of PMCs (pollen mother cells) could also help to determine the type of 2*n* gametes in males.

Unreduced gamete formation makes it possible to overcome post-zygotic barriers that occurred in interploid crosses due to endosperm abortion (Peloquin *et al.*, 1999). Most of the expected endosperm ploidy levels were observed in the different crosses [♀2*x* (*n*) × ♂4*x*, ♀2*x* (2*n*) × ♂4*x*] with the remarkable exception of the 2*x* × 2*x* cross with the formation of 2*n* ovules. In this latter case, the 3*x* individual resulted from plump seeds with a 3*x* endosperm. Histochemical analysis in *Dioscorea nipponica* by Torshilova *et al.* (2003) suggests that the embryo sac of Dioscoreacea would be of the monosporic octonuclear type, Polygonum, which leads to a 3*x* endosperm with a 2:1 ratio in a usual diploid cross. Therefore, the embryo sac of a diploid female that produces a 2*n* ovule is expected to contain two 2*n* polar nuclei. After fertilization by a normal male gamete, these two polar nuclei should normally generate a non-viable 5*x* endosperm with a 4:1 ratio. Since the observed endosperm ploidy was 3*x*, this suggests that only one of the two polar nuclei could have been fertilized. This phenomenon has been observed in other species such as in *Triticum aestivum* (You and Jensen, 1985). For triploid individuals resulting from a union between unreduced pollen and a reduced ovule, endosperms were shrivelled. Their expected endosperm ratio is 2:2, which, unfortunately, could not be measured in our experiment. This suggests that this type of gametic combination should not lead to the production of viable seeds as a consequence of endosperm dysfunction.

No triploid seedlings were obtained from the 1000 seeds from the intercytotype cross, 2*x* female × 4*x* male. Embryos were 3*x* but their endosperm were found to be 4*x*, consistent with the expected 2:2 maternal to paternal genome ratio, confirming that in *D. alata*, unbalanced endosperm leads to shrivelled seeds and unviable seeds. Since *in vitro* embryo rescue on such crosses gives rise to 3*x* hybrids, this scenario is confirmed.

In brief, since plump seeds from 2*n* ovules × *n* pollen have a better chance of germinating in the wild than shrivelled seeds from an *n* ovule × 2*n* pollen, the likely spontaneous triploid formation in *D. alata* is the cross of a diploid maternal gamete with a haploid paternal gamete (Fig. 5A).

Since 3*x* are highly infertile, the existence of fertile 4*x* plants remains to be documented. Theoretically, the two different ways that could have led to the production of the first tetraploids are USP (triploid bridge) and BSP (fusion of two unreduced gametes). Obtaining tetraploids via BSP has been demonstrated in several autopolyploid species such as red clover (Parrott *et al.*, 1985) and *Dactylis glomerata* (Bretagnolle and Lumaret, 1995). In the 2*x* × 2*x* progeny, no tetraploid was obtained by BSP on the 300 seeds examined. With an observed rate of 2*n* gamete formation of 1·3 %, the theoretical chance of obtaining a tetraploid

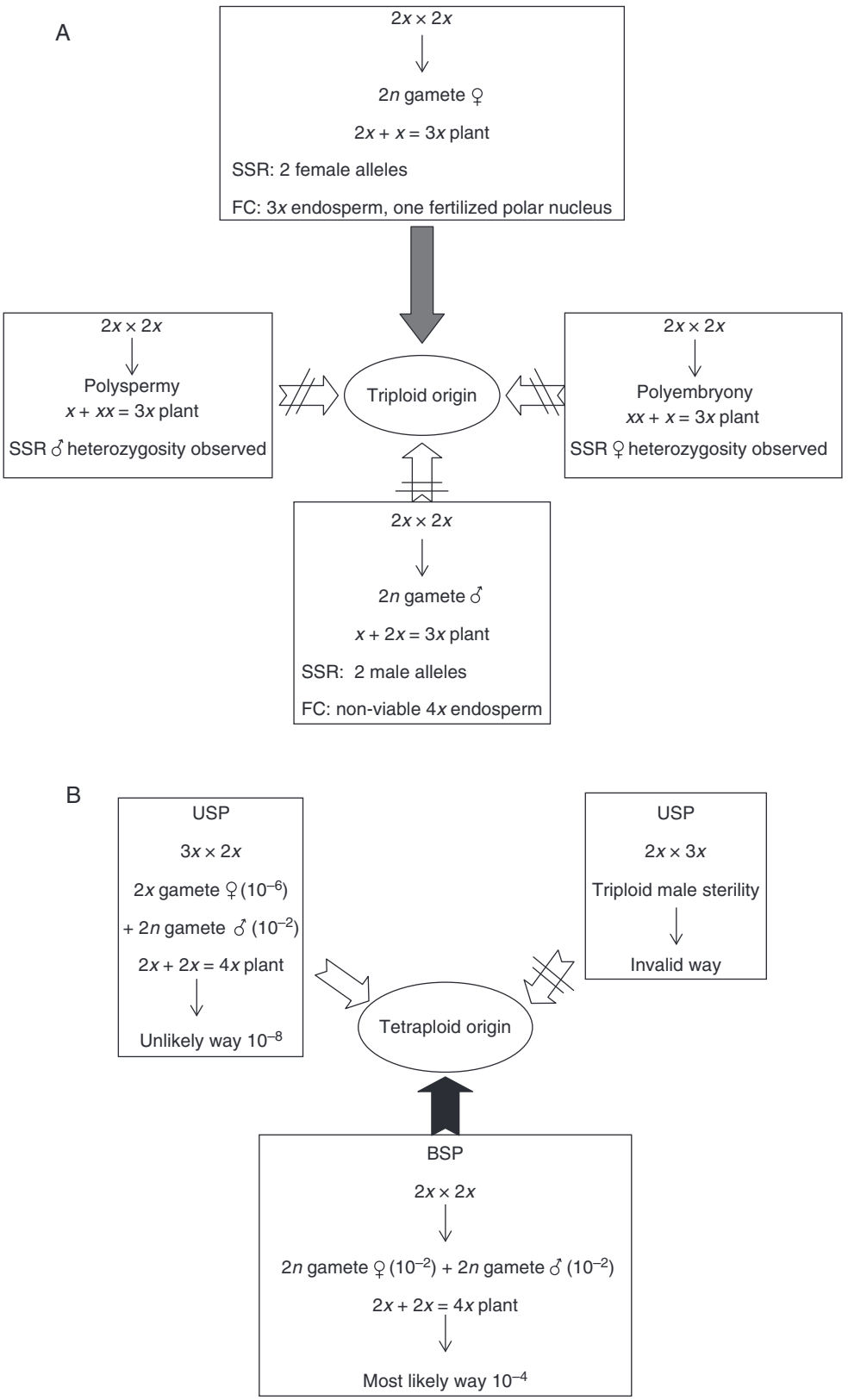


FIG. 5. Scheme representing the different pathways for the origins of *D. alata* polyploids. Triploid origin in (A) and tetraploid origin in (B). SSR, microsatellite analysis; FC, flow cytometry analysis; // indicates an invalid route.

would have been 1/2500. If this scenario could be verified by examining a much larger descent than in this study, it could be considered as much more likely than the triploid bridge for *D. alata* (Fig. 5B). Furthermore, it would normally not have a problem of seed development with the BSP scenario because a normal 6*x* endosperm with a 2:1 (4:2) ratio is expected.

No tetraploid seedlings were obtained by triploid bridges on a descent of 2000 seeds (i.e. a cross between triploids and diploids to obtain tetraploids). Our results confirm a general picture of a very high sterility rate of triploid females, as reported by Abraham and Nair (1991). These authors concluded that triploid females are sterile based on a study carried out on 27 triploid females crossed with a male diploid (50–130 pollinations).

In triploids, segregation of chromosomes at meiosis is complex because of the trivalent formation at metaphase I, which would lead to a majority of aneuploid gametes after the second meiotic division. The probability of obtaining viable euploid gametes (balanced) depends on the *x* chromosome number. Given that each chromosome has a 50 % chance of migrating to one of the cell's poles at anaphase I, the frequency of obtaining haploids, diploids and triploids is $1/2^x$, $1/2^x + 1/2^x$ and $1/2^x \times 3$, respectively. The chromosome base number of most species for which the triploid bridge has been demonstrated is quite low. *Cucumis sativus*, in which two tetraploids were obtained by USP on 545 seeds, has $x = 7$ pairs of chromosomes (Diao *et al.*, 2009).

It has been demonstrated that non-reduction gametic phenomena can increase the frequency of obtaining diploid and triploid gametes in triploids by SDR and FDR, respectively (Husband, 2004). This is the case for *C. angustifolium* ($x = 18$) where the high rate of unreduced gametes made it possible to obtain tetraploids. The rate of 2*n* gametes observed in *C. angustifolium* species is ten times higher than that reported by Ramsey and Schemske (1998), based largely on crop plants. Moreover, in this species that has an Oenothera-type monosporic embryo sac, no endosperm incompatibility phenomenon was observed in intercytotype crosses (Husband, 2004). In *D. alata*, whose chromosome base number is $x = 20$ (Arnau *et al.*, 2009), viable haploid, diploid and triploid gametes are predicted at frequencies of $1/2^{20}$, $1/2^{20} + 1/2^{20}$ and $1/2^{20} \times 3$, respectively, corresponding to very low rates for obtaining viable seedlings. Screening of triploid females in other collections would make it possible to check if USP (triploid bridge) is a possible formation mechanism of tetraploids. However, in the CIRAD *D. alata* germplasm collection, no gametic non-reduction has yet been observed for triploid females (unpublished data) via USP checking. In *D. alata*, in addition to a very low theoretical frequency for obtaining triploid balanced gametes (three out of 1 million), it would be assumed that endosperm incompatibility would lead to a heptaploid endosperm with a ratio of 6:1 (or tetraploid with a 3:1 ratio if only one polar nucleus is fertilized) and a likely non-viability for the seeds.

Moreover, our results revealed that in *D. alata*, as in other species (Burton and Husband, 2001; Hardy *et al.*, 2001), tetraploids can be obtained by intercytotype crosses ($2x \times 4x$) via gametic non-reduction in the diploid female progenitor. Thus, the gene flows between the diploid and tetraploid compartments may have contributed to enhancing diversity at the tetraploid level. Reciprocal crosses ($4x \times 2x$), although not carried out in this study, could produce tetraploids via gametic non-reduction in the male diploid progenitor. Once the tetraploid pool is established, mating could start between 4*x* individuals and lead to the recombination of diversity at this ploidy level.

Conclusions

The major conclusion of this study is that polyploids in *D. alata* might have appeared as a result of 2*n* gamete formation. The most likely origin of spontaneous triploids would be the union of an unreduced egg and a reduced pollen grain with normal 3*x* endosperm, whose formation is still unknown. Crossing a diploid female with a tetraploid male is a possible way to obtain triploids in *D. alata* but requires the use of embryo rescue, excluding the possibility that this mechanism could have contributed to enhancing the diversity of the wild triploid pool. Although tetraploids were not obtained in this experiment, probably since the sample of seeds germinated was too small, we could reasonably assume that their most likely primary origin in the wild would be BSP via the mating of two unreduced gametes produced by diploids. Gene flows between diploid and tetraploid compartments by intercytotype crosses ($2x \times 4x$) may have further contributed to broadening the allelic diversity at the tetraploid level, while recombination at the 4*x* level created new gene combinations. Screening of *D. alata* collections worldwide could make it possible to identify diploid progenitors with high 2*n* gamete production and to use them to enlarge the genetic diversity in the cultivated polyploid compartment. Knowledge of the origin of 2*n* gametes, SDR or FDR, will make it possible to increase the transmission pathway of the parental heterozygosity rate.

ACKNOWLEDGEMENTS

We would like to thank the CIRAD, the European Regional Development Fund (ERDF), the European Social Fund (ESF) and the Regional Council of Guadeloupe who financed this research.

LITERATURE CITED

Abraham K, Nair PG. 1991. Polyploidy and sterility in relation to sex in *Dioscorea alata* L. (Dioscoreaceae). *Genetica* **83**: 93–97.

Abraham K, Arnau G. 2009. *Use of DNA markers for genetically improving the productivity, palatability, storability and dry matter content of tubers of greater yam. Report of project no. 3000-B1*. New Delhi: Indo-French Society for the Promotion of Advanced Research (IFCPAR).

Andris M, Aradottir GI, Arnau G, *et al*. 2010. Permanent genetic resources added to Molecular Ecology Resources Database 1 June 2010–31 July 2010. *Molecular Ecology Resources* **10**: 1106–1108.

Arnau G, Maledon E, Bachand I, Abraham K. 2006. Production of interploid hybrids and molecular markers heterozygosity determination using microsatellite markers in the greater yam, *D. alata*: importance for the genetic improvement of the greater yam. In: *Proceedings of 14th Triennial Symposium of the International Society for Tropical Roots Crop*s. 20–26 November 2006, Thiruvananthapuram, Kerala, India.

Arnau G, Nemorin A, Maledon E, Abraham K. 2009. Revision of ploidy status of *Dioscorea alata* (Dioscoreaceae) by cytogenetic and microsatellite segregation analysis. *Theoretical and Applied Genetics* **118**: 1239–1249.

Arnau G, Abraham K, Sheela MN, Chair H, Sartie A, Asiedu R. 2010. Yams. In: Bradshaw JE. ed. *Root and tuber crops. Handbook of plant breeding*, Vol. 7. Berlin: Springer, 127–148.

Arnau G, Nemorin A, Maledon E, Nudol E. 2011. Advances on polyploid breeding in yam *D. alata*. *Proceedings of the First International*

Symposium on Roots, Rhizomes, Tubers, Plantains, Bananas and Papaya. 7–10 November 2011, Santa Clara, Cuba.

Ashley MV, Dow BD. 1994. The use of microsatellite analysis in population biology: background, methods and potential applications. *EXS* **69**: 185–201.

Ayensu ES, Coursey DG. 1972. Guinea yams: the botany, ethnobotany, use and possible future of yams in West Africa. *Economic Botany* **26**: 301–318.

Batygina TB, Vinogradova Gyu. 2007. Phenomenon of polyembryony. Genetic heterogeneity of seeds. *Russian Journal of Developmental Biology* **38**: 126–151.

Borg M, Brownfield L, Twell D. 2009. Male gametophyte development: a molecular perspective. *Journal of Experimental Botany* **60**: 1465–1478.

Burton TL, Husband BC. 2001. Fecundity and offspring ploidy in matings among diploid, triploid and tetraploid Chamerion angustifolium (Onagraceae): consequences for tetraploid establishment. *Heredity* **87**: 573–582.

Bretagnolle F, Lumaret R. 1995. Bilateral polyploidization in *Dactylis glomerata* L. subsp.*lusitanica*: occurrence, morphological and genetic characteristics of first polyploids. *Euphytica* **84**: 197–207.

Bretagnolle F, Thompson JD. 1995. Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants. *New Phytology* **129**: 1–22.

Carputo D, Barone A, Frusciante L. 2000. 2*n* gametes in the potato: essential ingredients for breeding and germplasm transfer. *Theoretical and Applied Genetics* **101**: 805–813.

Costa LM, Gutièrrez-Marcos JF, Dickinson HG. 2004. More than a yolk: the short life and complex times of the plant endosperm. *Trends in Plant Science* **9**: 507–514.

Diao WP, Bao SY, Jiang B, Cui L, Chen JF. 2009. Primary trisomics obtained from autotriploid by diploid reciprocal crosses in cucumber. *Sexual Plant Reproduction* **22**: 45–51.

Doležel J, Lucretti S, Schubert I. 1994. Plant chromosome analysis and sorting by flow cytometry. *Critical Reviews in Plant Sciences* **13**: 275–309.

Egesi CN, Asiedu R. 2002. Analysis of yam yields using the Additive Main Effects and Multiplicative Interaction (AMMI) model. *African Crop Science Journal* **10**: 195–201.

Gallais A. 2003. *Quantitative genetics and breeding methods in autopolyploid plants*. France: Editions Quae.

Gmitter FG, Ling XB, Deng XX. 1990. Induction of triploid *Citrus* plants from endosperm calli *in vitro*. *Theoretical and Applied Genetics* **80**: 785–790.

Hardy OJ, De Loose M, Vekemans X, Meerts P. 2001. Allozyme segregation and inter-cytotype reproductive barriers in the polyploid complex *Centaurea jacea*. *Heredity* **87**: 136–145.

Harlan JR, deWet JMJ. 1975. On O. Winge and a prayer: the origins of polyploidy. *Botanical Review* **41**: 361–390.

Husband BC. 2004. The role of triploid hybrids in the evolutionary dynamics of mixed-ploidy populations. *Biological Journal of the Linnean Society* **82**: 537–546.

Hutten RCB, Schippers MGM, Hermsen JGTh, Jacobsen E. 1995. Comparative performance of diploid and tetraploid progenies from 2*x*·2*x* crosses in potato. *Euphytica* **81**: 187–192.

Iwanaga M, Peloquin SJ. 1982. Origin and evolution of cultivated tetraploid potatoes via 2*n* gametes. *Theoretical and Applied Genetics* **61**: 161–169.

Köhler C, Scheid OM, Erilova A. 2010. The impact of the triploidblock on the origin and evolution of polyploid plants. *Trends in Genetics* **26**: 142–148.

Lange W, Wagenvoort M. 1973. Meiosis in triploid *Solanum tuberosum* L. *Euphytica* **22**: 8–18.

Lebot V. 2009. *Tropical root and tuber crops: cassava, sweet potato, yams and aroids*. Wallingford, UK: CABI Publishers.

Malapa R, Arnau G, Noyer JL, Lebot V. 2005. Genetic diversity of the greater yam (*Dioscorea alata* L.) and relatedness to *D. nummularia* Lam. and *D. transversa* Br. as revealed with AFLP markers. *Genetic Resources and Crop Evolution* **52**: 919–929.

Mendiburu AO, Peloquin SJ. 1977. Bilateral sexual polyploidization in potatoes. *Euphytica* **26**: 573–583.

Misuki I, Tani N, Ishida K, Tsumura Y. 2005. Development and characterization of microsatellite markers in a clonal plant, *Dioscorea japonica* Thunb. *Molecular Ecology Notes* **5**: 721–723.

Mizuochi H, Matsuzaki H, Moue T, Okazaki K. 2009. Diploid endosperm formation in *Tulipa* spp. and identification of a 1:1 maternal-to-paternal genome ratio in endosperms of *T. gesneriana* L. *Sexual Plant Reproduction* **22**: 27–36.

Mok DWS, Peloquin SJ. 1975. Three mechanisms of 2n pollen formation in diploid potatoes. *Journal Canadien de Génétique et de Cytologie* **17**: 217–225.

Muniyamma M. 1977. Triploid embryo from endosperm *in vivo*. *Annals of Botany* **41**: 1077–1079.

Navashin SG, Finn VV. 1951. On the history of the development of the chalazogamies *Juglans nigra* and *J. regia*. *Izbrannye trudy, Moscow-Leningrad* **1**: 271–325.

Nemorin A, Abraham K, David J, Arnau G. 2012. Inheritance pattern of tetraploid *Dioscorea alata* and evidence of double reduction using microsatellite marker segregation analysis. *Molecular Breeding* **30**: 1657–1667.

Ortiz R, Vuylsteke D. 1995. Factors influencing seed set in triploid *Musa* spp. L. and production of euploid hybrids. *Annals of Botany* **75**: 151–155.

Parisod C, Holderegger R, Brochmann C. 2010. Evolutionary consequences of autopolyploidy. *New Phytologist* **186**: 5–17.

Park SM, Wakana A, Kim JH, Jeong CS. 2002. Male and female fertility in triploid grapes (*Vitis* complex) with special reference to the production of aneuploid plants. *Vitis* **41**: 11–19.

Park TH, Kim JB, Hutten RCB, van Eck HJ, Jacobsen E, Visser RGF. 2007. Genetic positioning of centromeres using half-tetrad analysis in a 4*x*–2*x* cross population of potato. *Genetics* **176**: 85–94.

Parrott WA, Smith RR, Smith MM. 1985. Bilateral sexual tetraploidization in red clover. *Canadian Journal of Genetics and Cytology* **27**: 64–68.

Peloquin SJ, Boiteux LS, Carputo D. 1999. Meiotic mutants in potato: valuable variants. *Genetics* **153**: 1493–1499.

Peloquin SJ, Boiteux LS, Simon PW, Jansky SH. 2008. A chromosome-specific estimate of transmission of heterozygosity by 2*n* gametes in potato. *Journal of Heredity* **99**: 177–181.

Pichot C, El Maataoui M. 1997. Flow cytometric evidence for multiple ploidy levels in the endosperm of some gymnosperm species. *Theoretical and Applied Genetics* **94**: 865–870.

Ramachandran K. 1968. Cytological studies in Dioscoreaceae. *Cytologia* **33**: 401–410.

Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annual Review of Ecology and Systematics* **29**: 467–501.

Ramsey J. 2007. Unreduced gametes and neopolyploids in natural populations of *Achillea borealis* (Asteraceae). *Heredity* **98**: 143–150.

Scott RJ, Spielman M, Bailey J, Dickinson HG. 1998. Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* **125**: 3329–3341.

Seker M, Tuzcu O, Ollitrault P. 2003. Comparison of nuclear DNA content of citrus rootstock populations by flow cytometry analysis. *Plant Breeding* **122**: 169–172.

Shajeela PS, Mohan VR, Jesudas LL, Soris PT. 2011. Nutritional and antinutritional evaluation of wild yam. *Tropical and Subtropical Agroecosystems* **14**: 723–730.

Sliwinska E, Zielinska E, Jedrzejczyk I. 2005. Are seeds suitable for flow cytometric estimation of plant genome size? *Cytometry* **64**: 72–79.

Torshilova AA, Titova GE, Batygina TB. 2003. Female reproductive structures and seed development in *Dioscorea* Nipponicamakino (Dioscoreaceae). *Acta Biologica Cracoviensia Series Botanica* **45**: 149–154.

Tostain S, Scarcelli N, Brottier P, Marchand JL, Pham JL, Noyer JL. 2006. Development of DNA microsatellite markers in tropical yam (*Dioscorea* sp.). *Molecular Ecology Notes* **6**: 173–175.

Williams JH, Friedman WE. 2002. Identification of diploid endosperm in an early angiosperm lineage. *Nature* **415**: 522–526.

Yarmolyuk GI, Shiryaeva GI, Kulik AG. 1990. Endosperm embryony – a new form of apomixis in sugar beet. *Doklady Vsesoyuznoi Ordena Lenina i Ordena Trudovogo Krasnogo Znameni Akademii Sel'skokhozyaistvennykh Nauk im. V.I. Lenina* **3**: 8–11.

You R, Jensen WA. 1985. Ultrastructural observations of the mature megagametophyte and the fertilization in wheat (*Triticum aestivum*). *Canadian Journal of Botany* **63**: 163–178.

Publication 2

Arnau, G., Bhattacharjee, R., Chair, H., Malapa, R., Lebot, V., Abraham, K., Perrier, X., Petro, D., Penet, L., & Pavis, C. (2017). Understanding the genetic diversity and population structure of yam (*Dioscorea alata* L.) using microsatellite markers. *PloS One*, 12(3), e0174150.

PLOS ONE

RESEARCH ARTICLE

Understanding the genetic diversity and population structure of yam (*Dioscorea alata* L.) using microsatellite markers

Gemma Arnau[1]*, Ranjana Bhattacharjee[2]*, Sheela MN[3], Hana Chair[4], Roger Malapa[5†], Vincent Lebot[6], Abraham K[3], Xavier Perrier[4], Dalila Petro[7], Laurent Penet[7], Claudie Pavis[7]

1 Unité Mixte de Recherche Amélioration Génétique et Adaptation des Plantes (UMR Agap), Centre de Coopération International en Recherche Agronomique pour le Développement (CIRAD), Station de Roujol, Petit Bourg, Guadeloupe, France, **2** Bioscience Center, International Institute of Tropical Agriculture (IITA), PMB, Ibadan, Oyo State, Nigeria, **3** Central Tuber Crops Research Institute (CTCRI), Sreekariyam, Triruvananthapuram, India, **4** Unité Mixte de Recherche Amélioration Génétique et Adaptation des Plantes (UMR Agap), CIRAD, Montpellier, France, **5** Vanuatu Agricultural Research and Technical Centre (VARTC), Espiritu Santo PB, Vanuatu, **6** Unité Mixte de Recherche Amélioration Génétique et Adaptation des Plantes (UMR Agap), CIRAD, Port-Vila, Vanuatu, **7** ASTRO Agrosystèmes Tropicaux, INRA, Petit-Bourg (Guadeloupe), France

† Deceased.
* gemma.arnau@cirad.fr (GA); R.Bhattacharjee@cgiar.org (RB)

Citation: Arnau G, Bhattacharjee R, MN S, Chair H, Malapa R, Lebot V, et al. (2017) Understanding the genetic diversity and population structure of yam (*Dioscorea alata* L.) using microsatellite markers. PLoS ONE 12(3): e0174150. https://doi.org/10.1371/journal.pone.0174150

Editor: Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

Received: September 1, 2016

Accepted: March 3, 2017

Published: March 29, 2017

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded by grants from INRA [AIP Bio-Ressources: Div-Yam-2010] and by FEDER Guadeloupe [projects Caramba-Valexbiotrop].

Competing interests: The authors have declared that no competing interests exist.

Abstract

Yams (*Dioscorea* sp.) are staple food crops for millions of people in tropical and subtropical regions. *Dioscorea alata*, also known as greater yam, is one of the major cultivated species and most widely distributed throughout the tropics. Despite its economic and cultural importance, very little is known about its origin, diversity and genetics. As a consequence, breeding efforts for resistance to its main disease, anthracnose, have been fairly limited. The objective of this study was to contribute to the understanding of *D. alata* genetic diversity by genotyping 384 accessions from different geographical regions (South Pacific, Asia, Africa and the Caribbean), using 24 microsatellite markers. Diversity structuration was assessed via Principal Coordinate Analysis, UPGMA analysis and the Bayesian approach implemented in STRUCTURE. Our results revealed the existence of a wide genetic diversity and a significant structuring associated with geographic origin, ploidy levels and morpho-agronomic characteristics. Seventeen major groups of genetically close cultivars have been identified, including eleven groups of diploid cultivars, four groups of triploids and two groups of tetraploids. STRUCTURE revealed the existence of six populations in the diploid genetic pool and a few admixed cultivars. These results will be very useful for rationalizing *D. alata* genetic resources in breeding programs across different regions and for improving germplasm conservation methods.

Introduction

Yams (*Dioscorea* sp.) are important food security crops for millions of small-scale farmers in the tropical and subtropical regions of Africa, Asia, the Pacific, the Caribbean and Latin

America [1]. *Dioscorea alata* (known as the "greater yam" or the "winged yam") is one of the major cultivated species with wide geographical distribution [2]. It is currently second to *D. rotundata* in production volumes. Several traits of *D. alata* make it particularly valuable for commercial cultivation. These include high yield potential, ease of propagation, early growth vigour for weed suppression, and long storability of tubers [3, 4]. Tubers possess a high nutritional content with an average crude protein content of 7.4%, starch content of 75–84%, and vitamin C content ranging from 13.0 to 24.7 mg/100g [5].

Dioscorea alata is a dioecious species with a ploidy level ranging from 2n = 2x = 40 to 2n = 4x = 80 [6]. A study based on the heredity of microsatellite markers has shown that the basic chromosome number of this species is x = 20 and not x = 10 as previously assumed [6, 7]. This species was considered to be highly polyploid with six levels of ploidy (2n = 30, 40, 50, 60, 70 and 80) [8, 9]. However, it is now accepted that it has only three cytotypes (2n = 40, 60 and 80) and that the most common forms are diploids, followed by triploids and tetraploids are rare [6, 10, 11, 12].

Flowering of *D. alata* is erratic or absent in many cultivars [3, 13, 14, 15]. Cultivars have been exclusively clonally propagated by using small tubers or small pieces of tubers during hundreds or even thousands of years. Clonal propagation provides agronomical advantages but excludes sexual reproduction and could therefore represent a constraint for adaptation to biotic and abiotic stresses. It is also favorable to the spread of diseases, with pathogens allowed to adapt specifically to fixed genotype pools. The most serious disease in *D. alata* is anthracnose, which is caused by an airborne fungus *Colletotrichum gloesporioides* Penz. Anthracnose is found throughout the entire inter-tropical zone and can cause significant yield losses [16, 17, 18, 19, 20]. The importance of yams for food security has led to the establishment of several breeding programs for *D. alata*, in order to develop high-yielding cultivars with resistance to anthracnose, and tuber characteristics adapted to farmers' requirements [2, 21, 22, 23]. Nevertheless, the lack of knowledge on its origin and genetic diversity limits the efficacy of genetic improvement.

The center of origin of *D. alata* is not known. Based on archaeological evidence, it is thought to have been domesticated ca. 6000 years ago and is native to Asia-Pacific, but is not known in its wild state [12]. The greatest phenotypic variability in *D. alata* was observed in the southern part of Southeast Asia and in Melanesia, the probable center of origin for this species [15, 24, 25, 26, 27]. The South Pacific islands (Papua New Guinea, Fiji, New Caledonia, the Solomon and Vanuatu islands) have rich *ex situ* collections of *D. alata*, including more than 1000 cultivars [12].

A wide diversity also exists in India [28, 29] where several *ex situ* germplasm collections were established, including the most important collection at CTCRI (Central Tuber Crops Research Institute, Kerala, India) with 431 accessions. In addition, several international collections have been assembled, including those of the CRB-PT (Centre de Ressources Biologiques Plantes Tropicales INRA-CIRAD, Guadeloupe, France) and the IITA (International Institute of Tropical Agriculture, Ibadan, Nigeria), with 181 and 772 accessions of *D. alata*, respectively.

Lebot et al. [27] evaluated the diversity within 269 accessions of *D. alata* from different regions (South Pacific, Asia, Africa and the Caribbean) using enzymatic markers. However, the weak polymorphism of the four enzymatic systems did not reveal correlations between the groups of zymotypes and the geographic origins, ploidy levels and/or the phenotypic characteristics of the accessions.

Various molecular markers have been used to characterize the genetic diversity of the *D. alata* collections, including RAPDs [30], AFLPs [31] and SSRs [32, 33, 34]. SSR markers (microsatellites) are considered to be the markers of choice for analyzing genetic diversity because of their co-dominance, high reproducibility, high global mutation rates and polymorphism

[35, 36]. Nevertheless, the studies on *D. alata* involved a limited number of cultivars, and none was conducted at the global scale.

The aim of the present study was to analyze the genetic diversity and population structure of 384 *Dioscorea alata* accessions from different regions, including the South Pacific, Asia, Africa and the Caribbean using a common set of 24 microsatellite markers.

Material and methods

Plant material

Overall, 384 *D. alata* accessions originating from four collections were evaluated (S1 Table). These include two sets of germplasm: 363 landraces and 21 breeding lines, including 129 accessions from CRB-PT (Centre de Resources Biologiques Plantes Tropicales INRA-CIRAD, Guadeloupe), 90 from IITA (International Institute for Tropical Agriculture, Nigeria), 83 from CIRAD (Centre Internationale de Recherche Agronomique pour le developpement, Guadeloupe) and 82 from CTCRI (Central Tuber Crops Research Institute, India). The CIRAD collection is mainly composed of genotypes originating from Vanuatu (South Pacific). The accessions from IITA represented the core collection [37] developed from an entire collection of 772 *D. alata* West African landraces. The CRB-PT collection holds landraces from diverse geographical origins (Caribbean, South Pacific, South America). Overall, the germplasm was composed of 298 diploids, 51 triploids and 35 tetraploids. Ploidy levels of most accessions were determined in previous studies [6, 11, 31] except of 120 accessions (80 from IITA and 40 from CRB-PT), which were determined in this study using the protocol described in Arnau et al. [6].

Genotyping

Genomic DNA of the accessions was isolated in each institute using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) or the modified CTAB method as described by Sharma et al. [38].

Twenty-four SSR primer pairs (Table 1) developed from *D. abyssinica*, *D. praehensilis*, *D. japonica* and *D. alata*, were selected to analyze the accessions. Eleven primers were chosen based on their capacity to reveal high polymorphism and easy-to-score profiles (not stuttering) from a previous study (unpublished data, Arnau, 2013). The remaining thirteen were newly identified SSRs from *D. alata* [39]. Microsatellite alleles were scored using the software GeneMapper 4.0 (Applied Byosystems, USA).

PCR amplifications and gel electrophoresis were carried out on the GENTYANE genotyping platform (INRA UBP, UMR 1095, Clermont-Ferrand). Each locus was fluorescently labeled by M13 tail as described by Vallunen [40]. Four different fluorochroms were used (FAM^{TM}, $VIC^{R,}$ NED^{TM} and PET^{R}). The PCR amplifications were performed in a 10 μL final volume containing final concentrations of 1X AmpliTaq Gold 360 Master Mix (AB-life technologies), 0.05 μM labeled forward primer, 0.5 μM reverse primer and 25 ng of template DNA. For all loci the same PCR program was used, consisting of an initial denaturation at 95˚C for 10 min followed by a touchdown PCR consisting of 45 cycles with denaturation at 95˚C for 30 s; annealing for 30 s with temperature decreasing 1˚C every cycle from 62˚C to 56˚C (7 cycles), then 30 cycles at 55˚C and 8 cycles at 56˚C; and a final extension at 72˚C for 5 min. Two to four individual PCR products labeled with different fluorochroms were multiplexed and visualized using capillary gel electrophoresis on an ABI PRISM 3100 DNA sequencer (Applied Biosystems, Foster City, CA, USA).

Table 1. Genetic diversity detected in 367 D. alata accessions using 24 microsatellite markers. Ho, He and F*is* values were quantified only for diploids.

Origin [1]	SSR	EMBL[2]	Motif	Min.–max. size (bp)	Total alleles	Al < 1%[3]	Main allele frequency[4]	Ho[5]	He[6]	Fis[7]
D. A	Da3G04	AJ880369	(AC)12	282–306	9	1	0.80	0.83	0.86	0.03
D. A	Da1F08	AJ880368	(TG)13	161–185	9	3	0.88	0.32	0.39	0.19
D. A	Da2F10	[51]	(TG)14	108–151	14	2	0.46	0.65	0.67	0.04
D. A	Da1A01	AJ880381	(GT)8	201–222	7	1	0.85	0.52	0.51	-0.02
D. AB	Dab2D11	[51]	(TC)19	227–247	9	2	0.66	0.83	0.67	-0.25*
D. PR	Dpr3E10	[51]	(TCT)13(CTC)4	170–194	10	1	0.84	0.16	0.32	0.16*
D. PR	Dpr3B12	AJ880376	(TG)8	132–150	8	2	0.81	0.74	0.65	-0.15
D.J	DIJ034	AB201419	(AG)17	198–267	15	2	0.48	0.90	0.80	-0.14
D.J	DIJ443	AB201420	(AG)17	257–285	12	0	0.45	0.73	0.82	0.11*
D.J	DIJ0461	AB201423	(GA)16	120–144	8	3	0.64	0.78	0.71	-0.10
D.J	DIJ1045	AB201422	(TG)19	251–281	13	4	0.52	0.78	0.71	-0.10
D.A	mDaCIR2	FN677762	(CA)10	243–268	9	0	0.44	0.78	0.71	-0.10
D.A	mDaCIR11	FN677767	(AC)10	165–183	8	4	0.79	0.47	0.54	0.12*
D.A	mDaCIR13	FN677768	(GA)17	186–214	14	4	0.49	0.56	0.79	0.28*
D.A	mDaCIR17	FN677770	(AC)7	228–236	5	2	0.96	0.18	0.20	0.11*
D.A	mDaCIR20	FN677773	(GA)16	172–206	11	1	0.69	0.67	0.62	-0.08
D.A	mDaCIR26	FN677776	(TG)15(GA)14	171–219	12	2	0.59	0.60	0.74	0.18
D.A	mDaCIR57	FN677784	(TG)9	143–152	5	0	0.86	0.83	0.55	-0.28*
D.A	mDaCIR59	FN677786	(TC)11(CA)9	186–224	12	5	0.55	0.75	0.75	-0.01
D.A	mDaCIR60	FN677787	(CA)11	132–159	12	1	0.56	0.62	0.81	0.23
D.A	mDaCIR25	[38]	(AC)14	142–184	14	1	0.39	0.49	0.81	0.34*
D.A	mDaCIR61	FN677788	(AG)21	179–221	18	5	0.61	0.81	0.74	-0.09
D.A	mDaCIR63	[38]	(AG)12	155–180	8	1	0.68	0.53	0.65	0.18*
D.A	mDaCIR116	FN677800	(AG)8(AG)7	83–126	14	3	0.70	0.61	0.65	0.06*

[1]D. A, D. alata; D. AB, D. abyssinica; D. PR, D. praehensilis; D. J, D. japonica

[2] Registration number on EMBL database or publication reference

[3] Rare alleles with a frequency lower than 1%

[4] Highest frequency of an allele observed at this locus

[5] Observed heterozygosity.

[6] Expected heterozygosity

[7] Fixation index,*P <0.001

https://doi.org/10.1371/journal.pone.0174150.t001

Data analysis

The genotypic data was converted into a binary matrix that recorded the presence (1) or absence (0) of alleles for each microsatellite locus per accession. The following genetic parameters were estimated: number of alleles per locus, number of rare alleles with a frequency lower than 1% [41], highest frequency of an allele observed for each locus, gene diversity (expected heterozygosity, He), observed heterozygosity (Ho) and fixation index (Fis). The latest three parameters were quantified only for diploid accessions using the Genepop 4.0.10 software [42]. To test for deviation from Hardy-Weinberg proportion, the exact test was used (500 batches and 5000 iterations). These parameters were not calculated for polyploid cultivars because of the difficulty of unambiguously determining allele dosage in polyploids [43].

In order to compare accessions and their relationships, we calculated genetic distances between each pair of accessions with Dice dissimilarity coefficients [44], using the NTSYSpc software version 2.1 [45]. We then proceeded with two approaches: first with a Principal

coordinate analysis (PCoA) with the full study sample, and second a cluster analysis with UPGMA with a reduced sub-sample comprising all the diploid cultivars.

Indeed, since our study sample includes cultivars with three different ploidy levels (from diploids to tetraploids) whose parental contributions are unknown, we performed a Principal coordinate analysis (PCoA) to explore the diversity structure of all *D.alata* accessions. We used the Darwin software [46]. This method is indeed better than UPGMA to compare genetic diversity when plants with different ploidy levels are sampled [47].

We also carried out a cluster analysis for diploid accessions only using the UPGMA algorithm and the Dice similarity coefficient, using the NTSYSpc software version 2.1 [45]. Reliability and robustness of the clustering were based on 1000 random re-sampling conducted through the bootstrap procedure of TREECON software version 1.3. We also conducted a Mantel test [48] to assess the correlation between genetic distance and geographic origin of varieties for Vanuatu and Indian study varieties.

Population structure of diploids was also examined using a Bayesian approach using the software STRUCTURE V.2.3 [49]. This model assumes that the genome of individuals is a mixture of genes originating from K unknown ancestral populations. Under this model, the STRUCTURE algorithm estimated for each accession, the proportion of its genome (*q)* derived from the different clusters. So, individuals which may have mixed ancestry could be identified. In order to identify the number of populations (K) capturing the major structure in the data, we used the admixture model with a burn-in period of 30,000 steps and 10^6 MCMC (Markov Chain Monte Carlo) replicates. The number of clusters (K) evaluated ranged from 1 to 10. The analysis was performed using three independent runs for each simulated value of K. We then calculated ΔK [50], an ad-hoc statistics based on the rate of change in the log probability of data between successive K values, to have an estimation of the real number of clusters. A threshold of q = 0.80 was used to assign genotypes to one of inferred K clusters. The molecular variance among populations and accessions within the populations were calculated using an Analysis of Molecular Variance (AMOVA) approach using the software ARLEQUIN V.3.5 [51].

Results

Seventeen accessions presenting more than 25% of the missing data were removed from the analyses. The total number of accessions analyzed was 367, including 77 from India, 67 from Vanuatu, 89 from Caribbean, 87 from Africa, 13 from New Caledonia and 7 from French Guyana.

Genetic diversity

The analysis of 24 SSRs across 367 accessions allowed the identification of 256 alleles (Table 1), with the number of alleles per locus ranging from five (loci mDaCIR17 and mDaCIR57) to 18 (locus mDaCIR61) with an average of 10.7 alleles per locus. The number of rare alleles present in less than 1% of the sample varied from zero to five per marker. The frequency of the most common allele at each locus varied from 0.39 to 0.96 (Table 1).

Genetic diversity (expected heterozygosity) values ranged from 0.20 to 0.86 and the observed heterozygosity from 0.16 to 0.90. The mean genetic diversity of 0.66 indicated moderate to high levels of polymorphism in *D. alata*. Nine loci showed a significant heterozygote deficit while two loci presented significant excess. Overall, a slight heterozygote deficit was observed across all accessions (F_{is} = 0.03, $P<0.001$).

The Dice dissimilarity coefficients between all possible pairs of genotypes ranged from 0 and 0.86. Individuals with a distance of zero had the same alleles over the set of 24 loci and

Table 2. Details on duplicates from each collection based on genotypic profile across 24 SSR markers. The accessions grouped together presented identical allelic profiles at 24 SS loci.

Collection	Accession code	Geographic origin	Local name	Study code
CTCRI	Da322	[4]India	Unknown	212
CIRAD	VU579	[3]Vanuatu	Letslets Bokis	318
CIRAD	VU567	[3]Vanuatu	Letslets Bolos	323
CRB-PT	PT-IG-00040	[2]Puerto Rico	59_Vino white forme	151
CRB-PT	PT-IG-00052	[2]Puerto Rico	71_Smooth Statia	168
CRB-PT	PT-IG-00395	Unknown	452_Fafadro bis	173
IITA	TDa-1427	[1]Ghana	Alamun Gaga	54
IITA	TDa-1437	[1]Ghana	Adidianmawoba	66
CTCRI	Da40	[4]India	Elivalan	192
CTCRI	Da73	[4]India	Muramchari	223
CTCRI	Da28	[4]India	Kachil	194
CTCRI	Da39	[4]India	Poolakachil	253
CTCRI	Da143	[4]India	Gutu	233
CTCRI	Da78	[4]India	Kachil	247
CTCRI	Da95	[4]India	Kudakachil	234
CTCRI	Da22	[4]India	Chuvanna Maveran	199
CTCRI	Da100	[4]India	Parisakodan	220
CTCRI	Da70	[4]India	Thekkan Kachil	222
CTCRI	Da120	[4]India	Kaduvakkayyan	228
CTCRI	Da105	[4]India	Chenithakizhangu	261
CTCRI	Da48	[4]India	Vila Kachil	255
CTCRI	Da209	[4]India	Kachil	243
CRB-PT	PT-IG-00061	[2]Martinique	80_Igname d eau	183
CRB-PT	PT-IG-00030	[2]Martinique	48_67	186
CRB-PT	PT-IG-00045	[2]Martinique	64_St Vincent Violet	370
CRB-PT	PT-IG-00067	[3]New Caledonia	86_Wénéféla bis	374

[1]Africa
[2]Caribbean
[3]South Pacific
[4]Asia

https://doi.org/10.1371/journal.pone.0174150.t002

may be considered as possible duplicates. Several duplicates were detected (Table 2) in each of the collections. A set of six cultivars from three different germplasm collections [I-59, I-71, I-452 (CRB-PT), VU567, VU579 (CIRAD) and Da322 (CTCRI)] also showed identical genotypic profiles at 24 loci (Table 2).

Genetic structure

The population structure was first inferred by Principal coordinate analysis (PCoA) on all 376 accessions based on 256 alleles. The first two axes explained 57% of the genetic variability (Fig 1A and 1B).

Results showed that the accessions originating from India are primarily distributed in the lower part of the graph, whereas the accessions from Vanuatu are mainly distributed on the upper right-hand side (Fig 1A) revealing two distinct genepools. The Dice distances within the Indian collection ranged from 0 to 0.77, with an average value of 0.47 while that between

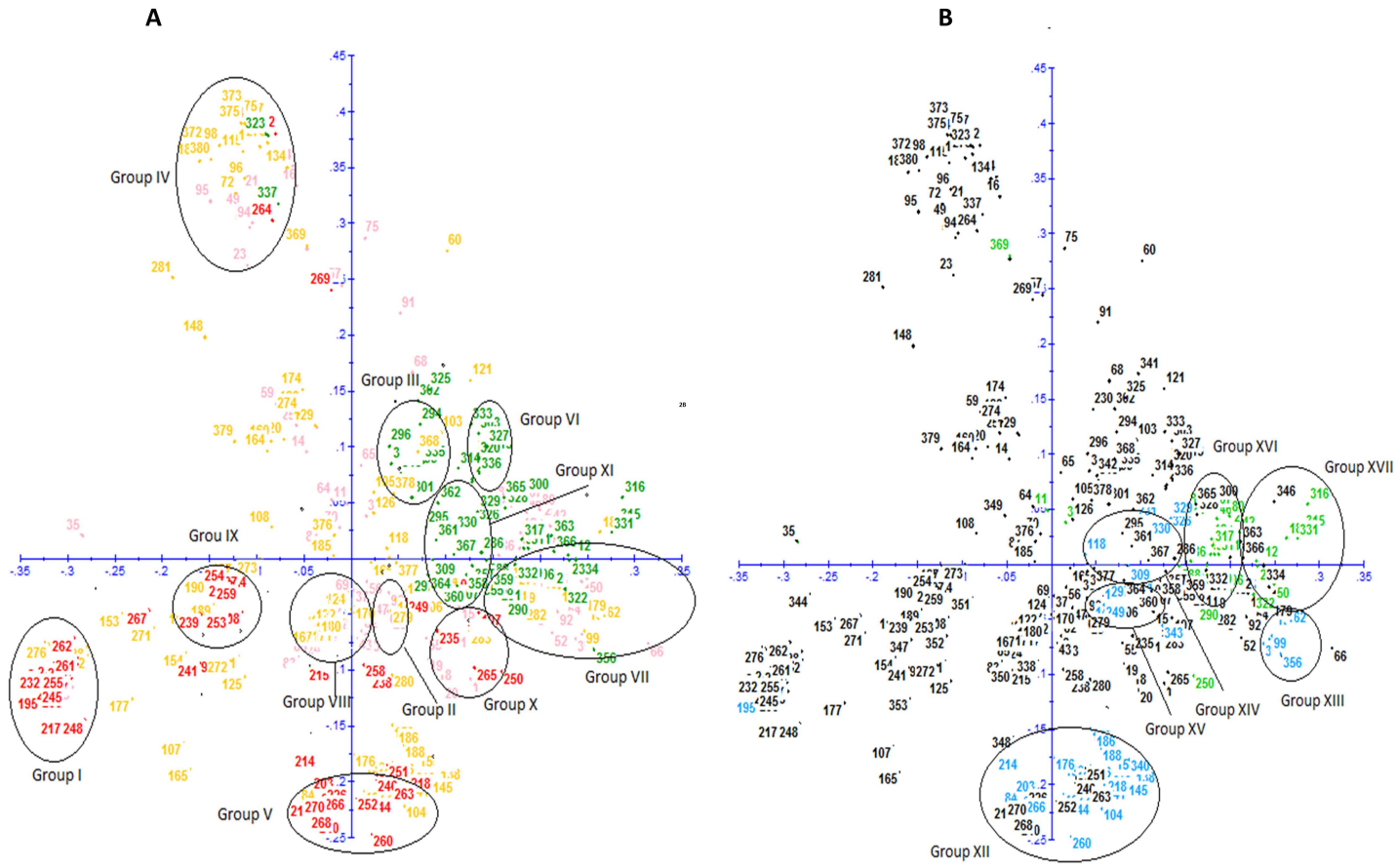


Fig 1A. Diagram showing the relationships among 367 accessions of *D. alata* based on Principal Coordinate Analysis (PCoA) using 24 microsatellites. Clones originated from Vanuatu are colored in green and those from CTCRI (India) in red. The accessions of the CRB-PT and the IITA are colored in orange and pink, respectively. **1B. Diagram showing the relationships among 83 polyploid accessions of *D. alata* based on Principal Coordinate Analysis(PCoA) using 24 microsatellites.** Tetraploids accessions are colored in green and triploids in blue.

https://doi.org/10.1371/journal.pone.0174150.g001

Vanuatu cultivars varied from 0 to 0.79, with an average of 0.49. The genetic distances between the Indian cultivars and those from Vanuatu ranged from 0.31 to 0.82, with an average value of 0.61. This value is significantly higher than the average value within each of the collections. Moreover, some of the alleles are specific to each of these two genepools. Twenty-nine alleles present in the Indian collection are absent from the Vanuatu genepool, including 12 rare alleles. Similarly, fifty-five alleles present in the Vanuatu genepool are absent in the Indian collection, including 15 rare alleles. Our Mantel test for these two pools demonstrated that correlation between genetics and origin was very high ($r = 0.408$, $P < 0.0001$).

The accessions of the international collections (IITA and CRB-PT) are distributed almost everywhere on the PCoA, showing that the diversity observed in India and in Vanuatu is well represented in CRB-PT and partially represented in IITA (Fig 1A). PCoA also showed that cultivars originating from New Caledonia (South Pacific) from CRP-PT collection are mostly distributed on the same part of the graph (side upper right-hand) as the Vanuatu's cultivars (South Pacific).

The 83 polyploid cultivars from the four collections (49 triploids and 34 tetraploids, S1 Table) are distributed in different zones of the PCoA graph and within proximity of the diploid cultivars (Fig 1B).

The combined analysis of PCoA and Dice distances allowed the identification of a total of seventeen different groups of cultivars composed of closely related cultivars (Fig 1A, Fig 1B and S1 Table) including eleven groups of diploids, four groups of triploids and two groups of tetraploids. The genetic distances between cultivars within each of these groups are lower than 0.25.

The UPGMA analysis supported the results of PCoA and clustered the diploids into the same eleven groups (Fig 2). Each group was supported with high bootstrap values ($\geq$ 89%), indicating high stability of relationships among the accessions. These groups thus reflected strong phylogenetic relatedness and can be seen as natural groupings based on the observed bootstrap values.

Diploid groups

Groups I, IV and VIII contained the highest number of cultivars (Fig 1A and Fig 2). Group I assembled 39 cultivars including 29 from India, eight from Caribbean and two from Africa. This group included the majority of duplicates identified within the CTCRI collection (Da40 and Da73, Da22 and Da95, Da70 and Da100, Da105 and Da120, Da78 and Da143, Da48 and Da209). Male cultivars cultivated in the Caribbean under the names of Pyramide, Brazo Fuerte and Brésil are also in this group. Group IV assembled 35 varieties from all geographic regions analysed (17 are from Caribbean, eleven from Africa, three from Vanuatu, two from India, one from New Caledonia, and one from French Guyana). This group included six accessions from three different collections that presented identical allelic profiles across 24 loci (I-59, I-71, I-452, VU567, VU579 and Da322), and can be considered as duplicates. It also listed two cultivars grown in the Caribbean with the names of Saint Vincent Blanc (I-54) and Saint Vincent Violet (I-64). The Dice distance between these two cultivars was 0.12 with 10 allelic differences at five loci. Group VIII represented 31 cultivars of which 20 are from Africa, 7 from Caribbean and 4 from French Guyana. This group included the cultivar, Pacala, which is cultivated and highly appreciated in the Caribbean but shows high susceptibility to anthracnose.

Group VII consisted of 19 cultivars of which 12 are from Africa, five from Caribbean and two from Vanuatu. The accessions from Africa (IITA cultivars TDa1427 and TDa1437) represented identical profiles across all 24 SSR loci (genetic distance equal to 0) and can be considered as duplicates. This group included the Florido variety selected by Martin and Rhodes [52] in Mayaguez, Puerto Rico, and distributed throughout the world. It also contains two cultivars that were classified in Vanuatu in the morphotypes group called Convar M3. This group is characterized by short oval tubers with white flesh, crumbly when cooked and highly appreciated in Vanuatu [15].

Group X assembled 15 cultivars including 9 from Africa, five from India and one from Caribbean. Group III included 12 cultivars, eight of which are from Vanuatu, one from New Caledonia and three from Caribbean. This group represents accessions known and cultivated in the Caribbean under the names of Kabusa and Lupias. The eight cultivars from Vanuatu were classified by Malapa [15] in the morphotype group Convar M1 since they present similarities of their aerial vegetative traits. These cultivars have elongate leaves with pointed foliar lobes, the base of their stem often has thorns, and their wings are not well developed on the stems. A wide variability exists for tuber shape, which may be long and cylindrical or compact and oval and the tuber flesh colour may be white or purplish. In addition, these cultivars produce male flowers and have a good tolerance to anthracnose in Vanuatu ([15]. Several of these cultivars were also evaluated in the Caribbean and presented good tolerance to anthracnose in Guadeloupe and in Martinique, including the accession (VU639a) known under the Vanuatu name of Malalagi [53].

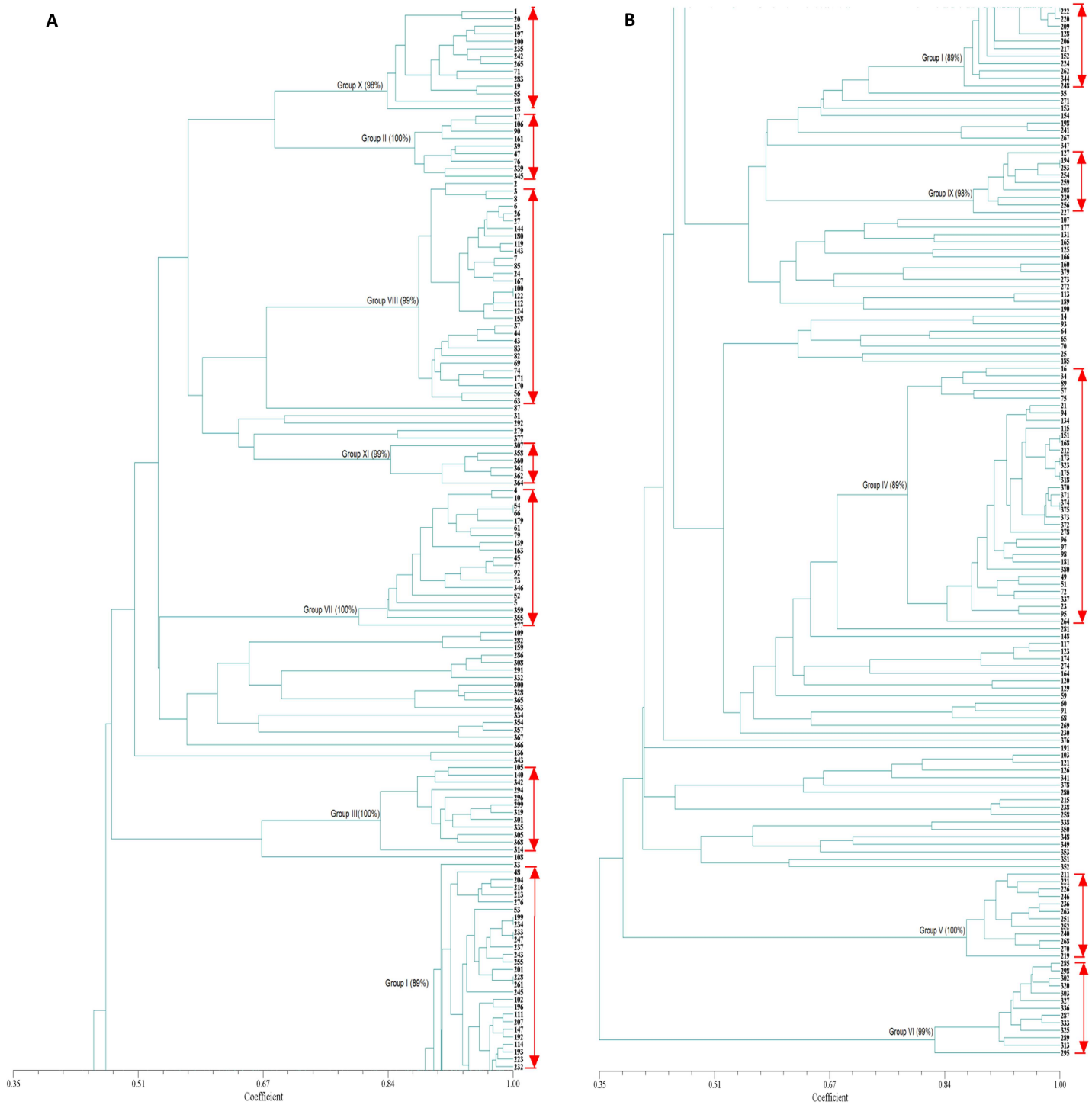


Fig 2. Dendrogram showing the relationships among 284 diploid accessions of *D. alata* based on UPGMA analysis using 24 microsatellites.

https://doi.org/10.1371/journal.pone.0174150.g002

Groups V and VI presented each 12 cultivars belonging to one collection. Group V is composed exclusively of cultivars from India and Group VI of cultivars from Vanuatu. Cultivars of group VI were classified in the morphotype convar M7 by Malapa [15]. They are characterized by hastate leaves whose colour are light green to yellow at maturity, by the presence of highly

developed and undulated wings on the stem, and have short tubers that are often ridged. These cultivars produce male flowers in an erratic manner. A wide variability exists for tuber flesh colour (white, purple, reddish) that is highly appreciated in Vanuatu for its elastic consistency when cooked. It has been observed that these cultivars have a good tolerance to anthracnose in Vanuatu [15]. Several of these cultivars were also evaluated in the Caribbean and had a good tolerance to anthracnose disease in Guadeloupe and in Haiti [53, 54]. The two closest cultivars (VU434a and VU487a) have only one allelic difference across all 24 SSR loci and represented a genetic distance of 0.01. These cultivars differ for their tuber flesh color, which is white in VU434a and purplish white in VU487a.

Group II contained 9 cultivars five of which were from Africa and four from the Caribbean region. This group is composed of cultivars known and cultivated in the Caribbean under the names of Kinabayo and Oriental that produce female flowers. Cultivar Oriental was resistant to anthracnose in Guadeloupe [11] and was used as a source of resistance to produce resistant hybrids [22]. Within this group, the genetic distance between two accessions bearing the same name from two different collections (Kinabayo from CRB-PT and Kinabayo from CIRAD) (0.15) was greater than the genetic distance between two cultivars with different names of CRB- PT collection (cultivars Oriental and Kinabayo), whose genetic distance was 0.08.

Group IX also assembled nine accessions of which eight are from India and one from Caribbean. Group XI assembled six cultivars from Vanuatu that were classified in the morphotype group, Convars M4 [15]. They are characterized by a wide variability in tuber shape which may be long, compact or irregular. In addition, these cultivars produce male flowers and presented a good tolerance to anthracnose in Vanuatu [15].

The remaining 82 diploids (61 cultivars and the 21 breeding lines) were either not grouped or clustered in small groups containing two to four cultivars.

Triploid groups

A total of four triploid groups were identified based on PCoA and genetic distances analysis (Fig 1B and S1 Table). The largest Group XII included 25 cultivars of which 18 are from Caribbean, six from India, two from French Guyana and 1 from New Caledonia. Two cultivars (I-80 and I-48) have identical profiles across 24 loci and can be considered as duplicates. This group includes three cultivars that share the name Tahiti (I-88: "French" Tahiti; I-87: "cultivated" Tahiti; and I-621: "snake" Tahiti). However, none of these cultivars are genetically identical across 24 SSR loci. They have nine to ten allelic differences (on four or five loci) and a genetic distance that varied between 0.10 and 0.11. Furthermore, cultivars belonging to this group have a characteristic allelic profile with three alleles at loci L31 (257 bp-269bp-275bp) and CIR59 (199bp-203bp-221bp), which is specific to this group and which made it possible to identify these cultivars.

Group XIII included six cultivars of which three are from New Caledonia, two from Africa and one from Vanuatu. This group includes accessions known in the Caribbean under the name of Goana. All cultivars belonging to this group have a characteristic allelic profile with three alleles at loci L31 (257 bp-261bp-265bp) and L30 (224bp-230bp-234bp), which is specific to this group. Group XV and XIV were the smallest and contained four and three accessions, respectively.

Tetraploid groups

Two tetraploid groups were identified based on PCoA and genetic distances analysis (Fig 1B and S1 Table). The largest assembled 19 cultivars (Group XVI) of which 13 are from Africa, five from Vanuatu and one from New Caledonia. These accessions were classified by Malapa [15] in the morphotype group convar M17. They are characterized by wide, cordated leaves

and rounded lobes, stems covered with thorns at their base, and developed and undulated wings. The tubers are long and either cylindrical or irregular in shape with purple flesh. They produce female flowers and showed tolerance to anthracnose in Vanuatu [15].

Group XVII included six cultivars of which one was from New Caledonia and five from Vanuatu. These cultivars were classified by Malapa [15] in the morphotype group convar M8. They produce male flowers and show tolerance to anthracnose in Vanuatu [15]. Several of these cultivars were also evaluated in the Caribbean and presented good tolerance to anthracnose disease in Guadeloupe and in Haiti [53, 54].

Bayesian analysis of population structure

The Bayesian model approach implemented in STRUCTURE V2.3.4 allowed the analysis of the population structure of diploids (284 accessions). The Evanno's method showed a peak of **Δk** for K = 6 supporting the presence of six genetically distinct populations (K = 6), here denoted as P1, P2, P3, P4, P5 and P6, respectively (Fig 3). Overall, 249 accessions (88%) were assigned to one of the six populations, where more than 80% of their inferred ancestry was derived from one of the model-based populations. P1 is composed of 115 accessions, among which 41 are from India, 50 from Caribbean, 20 from Africa and 4 from Vanuatu. This group is dominated by accessions originating from Indian genepool. P2 has 44 accessions, of which 28 are from Africa, seven from Caribbean, five from India and four from Vanuatu. This group corresponds to groups VII and X of PCoA. P3 included 20 accessions, among which 10 are from Caribbean and 10 from Vanuatu. P4 included 32 accessions, among which 20 are from Africa, 7 from Caribbean and 4 for French Guyana. These accessions corresponded to group VIII on PCoA. P5 included 25 accessions, among which three are from Caribbean, twelve from India, eight from Vanuatu and 1 from New Caledonia. These accessions corresponded to groups III and V on PCoA. P6 included 13 accessions and all originated from Vanuatu. These accessions are assigned to group VI on PCoA. The remaining 35 accessions (12%) included breeding lines and cultivars that showed admixed ancestry from different groups, including eight admixtures between P1 and P2, five between P2, P3 and P4, three between P1, P3 and P5, three between P2 and P4. The analysis of molecular variance (AMOVA) among populations indicated that 59.1% of the variation was due to differences within populations, while 40.9% was due to differences among populations (Table 3). The results of STRUCTURE analysis confirmed the existence of two divergent genepools in India and Vanuatu, and revealed some admixed cultivars, suggesting that these could have originated through hybridization between different populations. These results can be interpreted as population sets being clearly differentiated

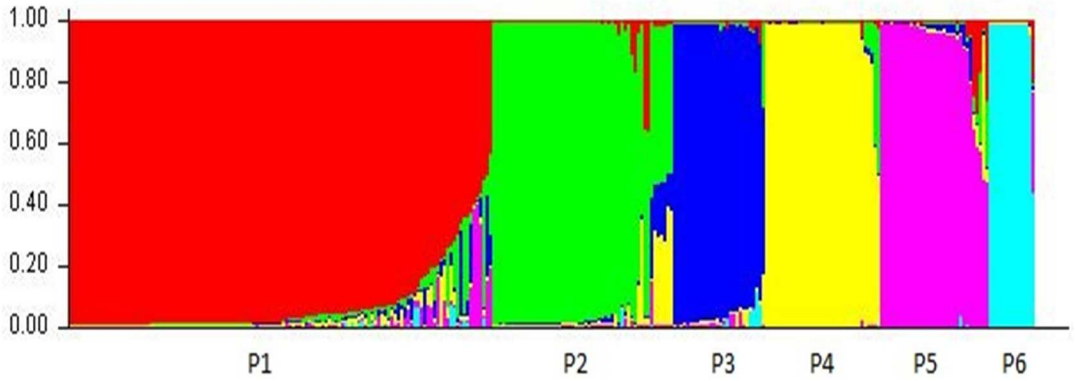


Fig 3. Structure of the genetic diversity of the 284 diploid accessions of *D. alata* at K = 6.

https://doi.org/10.1371/journal.pone.0174150.g003

Table 3. Analysis of molecular variance (AMOVA) among populations and within populations.

Source of variation	df	SS	CV	%Total	P value
Among populations	5	1031.7	8.9	40.9	<0.0001
Within populations	138	1770.6	12.8	59.1	<0.0001
Total	143	2802.3	21.7	100	

df, Degrees of freedom; SS, Sum of squares; CV, Variance components estimation
% Total, percentage of total variation.

https://doi.org/10.1371/journal.pone.0174150.t003

and with low historic admixture, to the exception of a few cultivars of recent origins. The groups defined by structure were thus possibly evolving and differentiating independently from each other with very low migration rates.

Discussion

This study represents the first approach that comprehensively investigates the genetic diversity within different large collections of *D. alata* using microsatellite markers. Several other studies on genetic diversity have been conducted on this species using different types of molecular markers including random amplified polymorphism DNAs [30, 55], amplified fragment length polymorphisms [15, 56, 57] and simple sequence repeat (SSR) markers [34, 58]. However, these included accessions from only one or two regions or collections. Lebot et al. [27] evaluated the genetic diversity of 269 accessions of *D. alata* originating from the South Pacific, Asia and the Caribbean using isozymes. However, due to low polymorphism of these markers, no relationships could be established between these accessions.

Microsatellite markers were effective for identifying polymorphism and for evaluating the genetic relationships between the 367 accessions analyzed. Regarding all 24 SSRs loci investigated, the estimated diversity index (H' = 0.66) was found to be high. Genetic diversity detected in this study is much higher than the one reported by Siqueira et al. [59] for 89 Brazilian *D. alata* cultivars (H' = 0.41). The diversity level detected in this species is also higher than what is reported in 146 *D. rotundata* cultivars (H' = 0.50) [60].

Our results revealed the existence of two divergent genepools in India and Vanuatu. In contrast to African species *D. rotundata*, whose center of origin is known and domestication is well documented [61, 62, 63], the center of origin for *D. alata* is actually not known yet. This species is supposed to have been domesticated in Southeast Asia [12] about 6000 years ago and then dispersed in India and the South Pacific Islands. Our results demonstrating a clear genetic differentiation between cultivars from India and Vanuatu are consistent with the existence of different secondary diversification centers in Asia and the South Pacific [12].

The introduction of *D. alata* in Africa and in New World takes place later in the 15th and 16th century. According to Prain and Burkill [64] this species was diffused from Asia to whole western Africa by the Portuguese in the 15th century. Degras [13] argued that it was introduced into New World at the beginning of 16th century with the slave trade.

PCoA and UPGMA analysis revealed a clear separation of the accessions into 17 major groups of genetically close cultivars, including eleven groups of diploids, four groups of triploids and two groups of tetraploids.

Two of the diploid groups assembled only cultivars from Vanuatu (VI and XI) and one group exclusively from India (V). The other diploid groups included cultivars from several geographic regions. Among these two included cultivars from India (I and X) and two from South Pacific (III and VII). One group (IV) assembled accessions from all geographic origins

analyzed (Caribbean, Africa, Vanuatu, India, New Caledonia and French Guyana). It included six cultivars from three different geographic origins (Caribbean, India and Vanuatu) which are genetically identical across 24 loci. The results indicate that farmers adopt good yam cultivars with superior attributes readily and these eventually get widely distributed.

The four triploid groups included cultivars from several geographic regions. Three groups consisted of cultivars from Vanuatu and one had cultivars from India. The biggest group (XII) included cultivars that have been largely diffused (India, Caribbean, French Guyana and New Caledonia). One group (XIV) included only cultivars from South Pacific (Vanuatu and New Caledonia). Group XIV assembled cultivars from Africa and South Pacific (Vanuatu and New Caledonia), while group XV included cultivars from Caribbean and Vanuatu.

Both tetraploid groups contained cultivars from several geographic regions, except from India. Group XVI included cultivars from Africa and South Pacific (Vanuatu and New Caledonia) and Group XVII assembled cultivars only from South Pacific (Vanuatu and New Caledonia).

Our results demonstrated a high degree of differentiation within this highly heterozygous species, probably due to limited gene flow, primarily restricted by the mating system of this species. This work showed also that *D. alata* cultivars originating from Africa and from the New World (Caribbean and South America) are genetically very similar to Asian and/or South Pacific cultivars.

The reproductive system is one of the important life-history characteristics that strongly influence genetic variability [65]. *Dioscorea alata* cultivars are exclusively vegetatively-propagated and present a low fertility. Two different hypotheses could explain this structuration in several groups of genetically close cultivars:

1. Accessions belonging to the same group are the outcome of former sexual events with the same parents or genetically related parents.
2. Accessions belonging to the same group are the outcome of the same initial clone that evolved via somatic mutations fixed by vegetative propagation.

Given the mutational rate of SSR markers (between 10^{-3} and 10^{-6}, depending on the species and on their position in the genome [36]) and the values of the genetic distances observed within these groups, it is likely that these two events contributed to the structuring of the genetic diversity in *D. alata*. The low number of allelic differences between cultivars assembled in group III, suggests the most likely hypothesis that this group was created from a single initial clone that evolved through mutations. Indeed, the number of allelic differences within this group corresponds to a total of 1 to 7 mutations, which is consistent with this hypothesis, on the basis of a mutation level of 10^{-3}–10^{-4}.

In contrast, the number of alleles differentiating the cultivars reassembled in other groups (ex: groups IV and VI) is higher to explain their origin from only an initial genotype. So, the most likely hypothesis is that these groups were created from several genetically related clones, which evolved through somatic mutations over time.

Our results are congruent with those obtained with DArTs to study the origin of *D. alata* cultivars in Vanuatu [66]. The low number of unique genotypes and the presence of numerous cultivars sharing a clonal origin generated a low varietal richness index (R = 0.26). This low index suggests that sexuality plays a minor role in the local diversification process. AFLPs have also highlighted the predominance of clonal reproduction in Vanuatu cultivars [15]. However, sexual recombination in *D. alata* appears to be rare in farmers' fields and makes the diversification of local cultivars by selecting spontaneous seedlings difficult. However, farmers can detect somaclonal variants and propagate them. The high genetic variability observed between *D.*

alata cultivars in Vanuatu is thought to be the results of multiple introductions of genetically distant individuals. The role of sexual reproduction and mutation in shaping the diversity has also been well documented for *D. rotundata* species [67].

Varietal diversity can be increased by controlled pollinations and breeding new cultivars and this has proven to be an effective way of genetic improvement [21, 22, 23, 68]. Most of the accessions included in our study are not fertile. Triploids are sterile [68, 69], as are several diploids and tetraploids. The results presented here allowed the identification of different groups of fertile cultivars with a good tolerance to anthracnose disease (Groups II, III, VI, XI, XVI and XVII) as well as several groups that contain cultivars known for the quality of their tubers: organoleptic, shape, dry matter content, and others (Groups III, VI, VII, VIII). These results will be useful for improving the germplasm management as well as for selecting genetically distant parents, even from different collections, to maximize allelic diversity and heterosis in breeding programs.

Much larger evaluations of *D. alata* genetic resources existing in Asian and South Pacific countries are still necessary in order to apprehend the genetic diversity available in these regions and allow their preservation and valorization in breeding programs.

Conclusions

Microsatellite markers and a worldwide sample material proved to be effective for identifying polymorphism and for evaluating genetic relationships between yam varieties, clarifying relationships between genetic diversity, geographic origins, and ploidy levels.

We did not identify any center of origin for this crop, but we demonstrated in our study the existence of two diversification pools, one from Vanuatu, and the other from India. High diversity levels were found in international germplasm collections, with CRB collection spanning throughout worldwide genetic diversity, while IITA collection was spanning a lesser extent but targeted more specific diversity component of variation for this species, possibly reflecting a finer sampling scheme.

Supporting information

S1 Table. Details of accessions with their accession code, geographical origin, local name, ploidy level and accession type included in the study. Identified groups based on PcoA and UPGMA analysis are also indicated.
(XLSX)

Acknowledgments

Gabrielle Clement and Amelie Amiot from Sherbrooke University (Canada) are thanked for their contribution to data analysis.

Author Contributions

Conceptualization: GA RB CP.

Formal analysis: GA.

Investigation: GA RB.

Methodology: RB CP DP SMN AK RM XP.

Writing – original draft: GA RB VL RM HC LP.

References

1. Ayensu ES, Coursey DG (1972) Guinea yams the botany, ethnobotany, use and possible future of yams in West Africa. Econ Bot 26: 301–318.
2. Abraham K, Nair PG (1990) Floral Biology and artificial pollination in *Dioscorea alata* L. Euphytica 48: 45–51.
3. Miege J (1952) L'importance économique des Ignames en Côte d'Ivoire. Répartition des cultures et principales variétés. Revue internationale de botanique appliquée et d'agriculture tropicale. 32e année, bulletin n°353–354, Mars-avril 1952. pp. 144–155.
4. Sartie A, Asiedu R (2014) Segregation of vegetative and reproductive traits associated with tuber yield and quality in water yam (*Dioscorea alata* L.). Afric J Biotech 13 (28): 2807–2818.
5. Muzac-Tucker I, Asemota HN, Ahmad MH (1993) Biochemical composition and storage of Jamaican yams (*Dioscorea* spp). J Sci Food Agric 62(3): 219–224.
6. Arnau G, Nemorin A, Maledon E, Abraham K (2009) Revision of ploidy status of *Dioscorea alata* L. (Dioscoreaceae) by cytogenetic and microsatellite segregation analysis. Theor Appl Genet 118: 1239–1249. https://doi.org/10.1007/s00122-009-0977-6 PMID: 19253018
7. Mignouna HD, Mank RA, Ellis THN, Van den Bosch N, Asiedu R, Abang MM et al (2002) A genetic linkage map of water yam (*Dioscorea alata* L) based on AFLP markers and QTL analysis for anthracnose resistance. Theo Appl Genet 105:726–735
8. Sharma AK, De DN (1956) Polyploidy in *Dioscorea*. Genetica 28: 112–120.
9. Hamon P, Brizard JP, Zoundjihekpon J, Duperray C, Borgel A (1992) Etude des index d'ADN de huit espèces d'ignames (*Dioscorea* sp.) par cytométrie en flux. Can J Bot 70: 996–1000.
10. Abraham K, Nair PG (1991) Polyploidy and sterility in relation to sex in *Dioscorea alata* L. (Dioscoreaceae). Genetica 83:93–97.
11. Gamiette F, Bakry F, Ano G (1999) Ploidy determination of some yam species (*Dioscorea* spp.) by flow cytometry and conventional chromosomes counting. Genet Res and Crop Evol 46: 12–27.
12. Lebot V (2009) Tropical root and tuber crops: cassava, sweet potato, yams and aroids. CABI, Wallingford, XIX 413pp.
13. Degras (1997) L'igname (Dioscorea spp) dans les sociétés des Amériques. In Proceedings, L'igname, plante séculaire et culture d'avenir, Montpellier, France, 33–44.
14. Abraham K (1997) Flowering deficiencies and sterility preventing natural seed set in *Dioscorea alata* cultivars. Trop Agric (Trinidad) Vol 74-n°4, 272–275.
15. Malapa R (2005) Description de la diversité de Dioscorea alata L. au Vanouatou à l'aide de marqueurs agromorphologiques et moléculaires (AFLP). Relations avec les autres espèces de la section Enanthiophyllum. PhD Thesis, Ecole Nationale Supérieure Agronomique de (Dioscorea rotundata Poir.) Rennes, France, 247pp.
16. Singh RD, Prasad N, Mathur RL (1966) On the taxonomy of the fungus causing anthracnose of *Dioscorea alata* L. Ind Phytopath 19: 65–71.
17. Fournet J, Degras L, Arnolin R, Jacqua G (1975) Essais relatifs à l'anthracnose de l'igname. Nouv. Agron. Antilles-Guyane. 1: 115–122.
18. Winch JE, Newhook FJ, Jackson GVH, Cole JS (1984) Studies of *Colletotrichum gloesporioides* disease on yam, *Dioscorea alata*, in Solomon Islands. Plant Path 33: 467–477.
19. Nwankiti AO, Ene LSO (1984) Advances in the study of anthracnose/bloch disease of Dioscorea alata in Nigeria. In: Proceedings of the 6th Symposium of the International Society of Tropical Root Crops, Lima, Peru. 633–640.
20. Abang MM, Winter S, Mignouna HD, Green KR, Asiedu R (2003) Molecular taxonomic, epidemiological and population genetic approaches to understanding yam anthracnose disease. Afr J Biotechnol 12: 486–496.
21. Egesi CN, Asiedu R (2002) Analysis of yam yields using the Additive Main Effects and Mutiplicative Interaction (AMMI) model. Afr Crop Sci J 10: 195–201.
22. Ano G, Anaïs G, Chidiac A (2002) Création et utilisation de variétés résistantes aux maladies, élements essentiels de la diversification agricole en Guadeloupe. Phytoma 551: 36–37.
23. Arnau G, Nemorin A, Maledon E, Nudol E (2011) Advances on polyploid breeding in yam D. alata. Proceding of the First International Symposium on Roots, Rhizomes, Tubers, Plantains, Bananas and Papaya. 7–10 November 2011 Santa Clara, Cuba, 4pp.
24. Bourret D (1973) Etude etnobotanique des Dioscoreacées alimentaires. Ignames de Nouvelle Calédonie. Thèse 3ème cycle. Faculté des Sciences, Université de Paris, France, 135pp.

25. Martin FM, Rhodes AM (1977) Intra-specific classification of *Dioscorea alata*. Trop Agric (Trinidad) 54: 1–13.
26. Sastrapadja S (1982) *Dioscorea alata*: its variation and importance in Java, Indonesia. In Miege J, Lyonga SN (Eds.), Yams. Ignames, pp. 45–49. Clarendon Press, Oxford, England.
27. Lebot V, Trilles B, Noyer JL, Modesto J (1998) Genetic relationships between *Dioscorea alata* L. cultivars. Genet Res and Crop Evol 45: 499–508.
28. Lakshmi KR, Easwari A (1980) Studies on variability and correlation in Asian greater yam (*Dioscorea alata* L.). J Root Crops 6: 29–32.
29. Velayudhan KC, Muralidharan VK, Amalraj VA, Thomas TA, Soudhamini P (1989) Studies on the morphotypic variability, distribution and genetic divergence in an indigenous collection of greater yam (*Dioscorea alata* L.) J Root Crops 15 (2): 79–89.
30. Asemota HN, Ramser J, Lopez-Peralta C, Weising K, Kahl G (1996) Genetic variation and cultivar identification of Jamaican yam germplasm by random amplified polymorphic DNA analysis. Euphytica 92: 341–351.
31. Malapa R, Arnau G, Noyer JL, Lebot V (2005) Genetic diversity of the greater yam (*Dioscorea alata* L.) and relatedness to *D. nummularia* Lam. and *D. transversa* Br as revealed with AFLP markers. Genet Resour Crop Evol 52: 919–929.
32. Siqueira MV, Marconi TG, Bonatelli ML, Zucchi MI, Veasey EA (2011) New microsatellite loci for water yam (*Dioscorea alata*, Dioscoreaceae) and cross-amplification for other *Dioscorea* species. Am J Bot 98 (6): 144–146.
33. Sartie A, Asiedu R, Franco J (2012) Genetic and phenotypic diversity in a germplasm working collection of cultivated tropical yams (*Dioscorea* spp.). Gen Res Crop Evol 59: 1753–1765.
34. Otoo E, Anokye ML, Asare PA, Telleh JP (2015) Molecuar categorization of some water yam (*Dioscorea alata* L.) gerplasm in Ghana using microsatellites (SSR) markers. J Agric Sci 7 (10): 226–238.
35. Ellegren H (2002) Microsatellite evolution: a battle between replication slippage and point mutation. Trends in Genet 18: 70–75.
36. Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC et al (2002) Rate and pattern of mutation at microsatellite loci in maize. Molec Biol Evol 19: 1251–1260. PMID: 12140237
37. Mahalaskshmi V, Ng Q, Atalobhor J, Ogunsola D, Lawson M, Ortiz R (2007) Development of a West African yam *Dioscorea* spp. core collection. Genet Resour Crop Evol 54: 1817–1825.
38. Sharma K, Mishra AK, Misra RS (2008) A simple and efficient method for extraction of genomic DNA from tropical tuber crops. Afr J Biotech 7:1018–1022.
39. Andris M, Aradottir GI, Arnau G, Audzijonyte A, Bess EC, Bodadonna F, et al. (2010) Permanent Genetic Resources added to Molecular Ecology Resources Database 1 June 2010–31 July 2010. Mol Ecol Resour 10: 1106–1108. https://doi.org/10.1111/j.1755-0998.2010.02916.x PMID: 21565125
40. Vallunen JA (2006) Protocols for Working with M13-Tailed Primers (v. 2.3.1).
41. Kimura M (1983) Rare variant alleles in the light of the neutral theory. Mol Biol Evol 1 (1): 84–93. PMID: 6599962
42. Raymond M, Rousset F (1995) GENEPOP (version 1, 2): population genetics software for exact tests and eucumenicism. J Hered 86: 248–249.
43. Pfeiffer T, Roschanski AM, Pannell JR, Korbecka G, Schnittler M (2011) Characterization of microsatellite loci and reliable genotyping in a polyploid plant, *Mercurialis perennis* (Euphorbiaceae). J Hered 102: 479–488. https://doi.org/10.1093/jhered/esr024 PMID: 21576288
44. Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26: 297–302.
45. Rohlf FJ (2000) NTSYSpc.: Numerical taxonomy and multivariate analysis system. Version 2.1. Exeter Software, Setauket, NY.
46. Perrier X, Jacquemoud-Collet JP (2006) DARwin software Version 5.0.155. CIRAD: http://darwin.cirad.fr/darwin.
47. Kloda JM, Dean PDG, Maddren C, MacDonald DW, Mayes S (2008) Using principle component analysis to compare genetic diversity across polyploidy levels with plant complexes: an example from British Restharrows (*Ononis spinosa* and *Ononis repens*). Hered 100: 253–260.
48. Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27: 311–335.
49. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959. PMID: 10835412

50. Evanno G, Regnault S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x PMID: 15969739

51. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evol Bioinform Online 1: 47–50.

52. Martin FW, Rhodes AM (1973) Correlations among greater yam (*Dioscorea alata* L.) cultivars. Tropic Agric (Trinidad) 50: 1–13.

53. Arnau G, Maledon E, Nudol E, Denis A (2009) Appui au développement de la filière igname: élargissement de la gamme variétale. Report Project PO-PDR-Feader-Odeadom, 20 p.

54. Arnau G, Maledon E, Prophete E, Scutt R, Joseph CD (2014) Varietés d'igname introduites à Haïti. Report Project Interreg Caraïbes DEVAG- Reseau Caribeen pour le développement de systèmes horticoles agroécologiques, 12 p.

55. Zannou A, Agbicodo E, Zoundjihekpon J, Struik PC, Ahanchede A, Kossou DK et al (2009) Genetic variability in yam cultivars from the Guinea-Sudan zone of Benin assessed by random amplified polymorphic DNA. Afr J Biotechnol 8: 26–36.

56. Egesi CN, Asiedu R, Ude G, Ogunyemi S, Egunjobi JK (2006) AFLP marker diversity in water yam (*Dioscorea alata* L.). Plant Genet Resour 4:181–187.

57. Tamiru M, Becker HC, Maass BL (2007) Genetic diversity in yam germplasm from Ethiopia and their relatedness to the main cultivated *Dioscorea* species assessed by AFLP markers. Crop Sci 47: 1744–1753.

58. Siqueira MV, Dequigiovanni G, Corazon-Guivin M, Feltran J, Veasey E (2012) DNA fingerprinting of water yam (*Dioscorea alata* L.) cultivars in Brazil based on microsatellite *markers*. Hort Brasil 30: 653–659.

59. Siqueira MV, Bonatelli ML, Gunther T, Gawenda I, Schmid KL, Pavinato VAC et al (2014) Water yam (*Dioscorea alata* L.) diversity pattern in Brazil: an analysis with SSR and morphological markers. Genet Resour Crop Evol 61: 611–624.

60. Tostain S, Agbangla C, Scarcelli N, Mariac C, Dainou O, Berthaud J et al (2007) Genetic diversity analysis of yam cultivars (*Dioscorea rotundata* Poir.) in Benin using simple sequence repeat (SSR) markers. Plant Genet Resour 5: 71–81.

61. Vernier P, Orkwor GC, Dossou AR (2003) Studies on yam domestication and farmers' practices in Benin and Nigeria. Outlook on Agriculture 32, 35_41.

62. Scarcelli N,Tostain S, Mariac C et al. (2006) Genetic nature of yams (D*ioscorea* sp.) domesticated by farmers in Benin (West Africa). Genet Resource and Crop Evolution 53, 121–130.

63. Chair H, Cornet D, Deu M, Baco N, Agbangla, Duval MF et al (2010) Impact of farmer selection on yam genetic diversity. Cons Genet 11(6): 2255–2265.

64. Prain D, Burkill IH (1939) An account of the genus *Dioscorea*. 2. Species which turn to the right. Ann. Roy. Bot. Garden Calcutta 14(2): 211–528.

65. Clegg MT, Epperson BK, Brown AHD (1992) Genetic diversity and reproductive system. In: Dattée Y, Dumas C, Gallais A (eds) Reproductive biology and plant breeding. Springer, Berlin, pp 311–323.

66. Vandenbroucke H, Mournet P, Vignes H, Chair H, Malapa R, Duval M-F et al (2015) Somaclonal variants of taro (*Colocasia esculenta* Schott) and yam (*Dioscorea alata* L.) are incorporated into farmers' varietal portfolios in Vanuatu. Genet Resour Crop Evol 63: 495–511.

67. Scarcelli N, Tostain S, Vigouroux Y, Luong V, Baco MN, Agbangla C et al (2011) Genetic structure of farmer-managed varieties in clonally-propagated crops. Genetica 139: 1055–1064. https://doi.org/10.1007/s10709-011-9607-8 PMID: 21898046

68. Abraham K, Nair PG, Sreekumari MT, Unnikrishnan (1986) Seed set and seedling variation in greater yam (*Dioscorea alata* L.). Euphytica 35: 337–342.

69. Némorin A, David J, Maledon E, Nudol E, Dalon J, Arnau G (2013) Microsatellite and flow cytometry analysis to help understand the origin of *Dioscorea alata* polyploids. Annals of Botany 112 (5): 811–819. https://doi.org/10.1093/aob/mct145 PMID: 23912697

Publication 3

Cormier, F., Lawac, F., Maledon, E., Gravillon, M. C., Nudol, E., Mournet, P., Vignes, H., Chair, H., & Arnau, G. (2019). A reference high-density genetic map of greater yam (*Dioscorea alata* L.). *Theoretical and Applied Genetics*, 132(6), 1733-1744.

Theoretical and Applied Genetics (2019) 132:1733–1744
https://doi.org/10.1007/s00122-019-03311-6

ORIGINAL ARTICLE

A reference high-density genetic map of greater yam (*Dioscorea alata* L.)

Fabien Cormier[1,2] · Floriane Lawac[1,2,3] · Erick Maledon[1,2] · Marie-Claire Gravillon[1,2] · Elie Nudol[1,2] · Pierre Mournet[1,4] · Hélène Vignes[1,4] · Hâna Chaïr[1,4] · Gemma Arnau[1,2]

Received: 6 November 2018 / Accepted: 11 February 2019 / Published online: 20 February 2019

Abstract

Key message **This study generated the first high-density genetic map for *D. alata* based on genotyping-by-sequencing and provides new insight on sex determination in yam.**

Abstract Greater yam (*Dioscorea alata* L.) is a major staple food in tropical and subtropical areas. This study aimed to produce the first reference genetic map of this dioecious species using genotyping-by-sequencing. In this high-density map combining information of two F1 outcrossed populations, 20 linkage groups were resolved as expected and 1579 polymorphic markers were ordered. The consensus map length was 2613.5 cM with an average SNP interval of 1.68 cM. An XX/XY sex determination system was identified on LG6 via the study of sex ratio, homology of parental linkage groups and the identification of a major QTL for sex determination. Homology with the sequenced *D. rotundata* is described, and the median physical distance between SNPs was estimated at 139.1 kb. The effects of segregation distortion and the presence of heteromorphic sex chromosomes are discussed. This *D. alata* linkage map associated with the available genomic resources will facilitate quantitative trait mapping, marker-assisted selection and evolutionary studies in the important yet scarcely studied yam species.

Introduction

Yams (*Dioscorea* spp.) are important food security crops that are grown in tropical and subtropical regions (Coursey 1967). They are dioecious herbaceous vines cultivated for their starchy tubers with a high nutritional content (Muzac-Tucker et al. 1993). *D. rotundata* and *D. alata* are the two main cultivated species (Ayensu and Coursey 1972) and belong to the same botanical section, i.e., Enantiophyllum (Wilkin et al. 2005), which is one of the latest diverging lineages in *Discorea* (Viruel et al. 2016).

Greater yam (*D. alata* L.) also named water or winged yam ranks second in production, and it is the most widely distributed yam species in the world (Abraham and Gopinathan Nair 1990). It is a strictly dioecious and polyploid species ($2n = 40, 60, 80$) with a basic chromosome number of 20 (Arnau et al. 2009). Diversity studies have shown that the most common forms are diploids, followed by triploids, and that tetraploids are rare (Arnau et al. 2017). It is superior to most cultivated yam species in terms of yield potential (especially under low soil fertility), ease of propagation, competition with weeds (early vigor) and tuber storability (Sartie and Asiedu 2014). Consequently, the importance of *D. alata* in terms of food security has given rise to several genetic improvement programs throughout tropical regions which are aimed at developing new varieties with high yield, tuber quality and resistance to pests and diseases (Abraham and Gopinathan Nair 1990; Egesi and Asiedu 2002; Arnau et al. 2011) such as anthracnose, caused by *Colletotrichum gloeosporioides* (Abang et al. 2004). Breeding of this

Communicated by Matthew N Nelson.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00122-019-03311-6) contains supplementary material, which is available to authorized users.

✉ Fabien Cormier
fabien.cormier@cirad.fr

1 CIRAD, UMR AGAP, 97170 Petit-Bourg, Guadeloupe, France

2 Univ. Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France

3 VARTC, P.O. Box 231, Luganville, Santo, Vanuatu

4 CIRAD, UMR AGAP, 34398 Montpellier, France

heterozygous crop is essentially carried out on the basis of phenotypic observations and is a long and difficult process.

One prerequisite for the development of marker-assisted breeding tools is the development of linkage analysis or/and association mapping. Both approaches are based on ordered genetic information. However, the *D. alata* genome has yet to be completely sequenced, and it is only available as unordered scaffolds (*D. alata* genome assembly Version 1, Water Yam Genome Project, ftp://yambase.org/genomes/Dioscorea_alata/). Two linkage maps were constructed using dominant amplified fragment length polymorphism markers (AFLP; Mignouna et al. 2002; Petro et al. 2011).

Codominant markers such as microsatellites (simple sequence repeats, SSRs) and single nucleotide polymorphisms (SNPs) are choice markers for plant breeding applications as they allow estimation of additive and dominant allelic effects. SSR markers have been generated in *D. alata* and other *Dioscorea* spp. (Terauchi and Konuma 1994; Misuki et al. 2005; Tostain et al. 2006; Hochu et al. 2006; Andris et al. 2010; Saski et al. 2015). Moreover, an EST-SSR genetic linkage map containing 380 markers was recently published (Bhattacharjee et al. 2018). Thus, the marker number is still limited for genome-wide approaches necessary for association genetics, and their implementation cost is relatively high.

With the development of next-generation sequencing (NGS) methods, SNPs are now the most widely available markers for high-throughput genotyping. Genotyping-by-sequencing (GBS) allows the detection and genotyping of tens of thousands of SNPs in many individuals (DePristo et al. 2011; Davey et al. 2011; Elshire et al. 2011), resulting in an unparalleled cost per data point when screening for codominant polymorphisms in large panels and for constructing highly saturated genetic maps (Poland et al. 2012, Ward et al. 2013).

The objective of the present study was to overcome the main limitations in identifying genomic regions linked to agronomic traits of interest in *D. alata* by establishing a high-density genetic map of *D. alata* using genotyping-by-sequencing. The relationship between genetic mapping and sex determination was also investigated since dioecy can be related to the chromosome architecture (Kumar et al. 2014).

Materials and methods

Materials

The mapping populations consisted of two greater yam (*Dioscorea alata*) full-sib F1 segregating populations. The hybridizations were performed in the French West Indies (Roujol, Petit-Bourg, Guadeloupe) using diploid parents (flow cytometry; Arnau et al. 2009). Population A was derived from a cross between a female breeding line (74F) developed at the French Agricultural Research Centre for International Development (CIRAD) and a male Caribbean landrace (Kabusa). Population B was derived from a cross between the same female (74F) and another male breeding line developed at CIRAD (14 M).

Overall, 250 and 360 pollinations were manually carried out for population A and population B, respectively. Once harvested (60–70 days after pollination), the fruits were left to soak in 70% isopropyl alcohol and then in 12% sodium hypochlorite for 5 min in each solution before rinsing with distilled water. Embryo rescue procedures were performed to conserve clean material and speed up multiplication. A total of 140 and 280 individuals were micropropagated for population A and population B, respectively. All progenies and parents were then transferred to an experimental field in two complete blocks with nine repetitions.

DNA extraction and genotyping-by-sequencing (GBS)

Young leaves from the same vine for each progeny and parent were collected, stored in coffee filters and then dried at 45 °C overnight. Total genomic DNA extractions were performed from dried leaves by an automated method adapted from Risterucci et al. (2009) on Biomek FXP (Beckman Coulter, CA, USA) and using the NucleoMag Plant Kit (Macherey–Nagel, Germany). DNA samples were quantified with a Fluoroskan Ascent FL fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Genomic DNA quality was checked using agarose gel electrophoresis. A genomic library was prepared using PstI-MseI (New England Biolabs, Hitchin, UK) restriction enzymes with a normalized 200 ng quantity of DNA per sample. The procedures published by Elshire et al. (2011) were followed; however, the common adapter was replaced to be complementary to MseI recognition site. Digestion and ligation reactions were conducted in the same plate. Digestion was conducted at 37 °C for 2 h and then at 65 °C for 20 min to inactivate the enzymes. The ligation reaction was done using T4 DNA ligase enzymes (New England Biolabs, Hitchin, UK) at 22 °C for 1 h, and the ligase was then inactivated by heating at 65 °C for 20 min. Parents were replicated twice per plate. Ligated samples were pooled and PCR-amplified (18 cycles). The PCR-amplified libraries were purified using the Wizard PCR preps DNA purification system Promega (Madison, USA) and verified with the Agilent D5000 ScreenTape (Santa Clara, USA). Single-end sequencing of 150 base-pair reads was performed in a single lane on an Illumina HiSeq 3000 system (at the GeT-PlaGe platform in Toulouse, France). Losses during micropropagation and transfer to the field led to the genotyping-by-sequencing of 121 progenies for population A (74F × Kabusa) and 193 for population B

(74F × 14 M). Parents were replicated to ensure SNP detection and high-quality parental information for the estimation of marker segregation types.

SNP calling and filtering

Raw sequencing data were demultiplexed with GBSX v1.2 (Herten et al. 2015). Cutadapt v1.9 (Martin 2011) was used to trim adapters (options: -a AGATCGGAAGAGCG -O 10 -q 20,20 -m 30). SNP calling was done using the process_reseq. 1.0.py (python2) program followed by site pre-filtering using the VcfPreFilter.1.0.py (python2) program with the default parameters. Both programs are part of the VcfHunter package (Garsmeur et al. 2018; available at https://github.com/SouthGreenPlatform/VcfHunter/). As there was no complete *D. alata* reference genome, reads were aligned to the *D. rotundata* reference genome (pseudo-chromosomes BDMI0100001-21; Tamiru et al. 2017). SNPs were thus named according to their position in this *D. rotundata* reference genome.

SNPs and progenies were filtered by population using the following filters: minimum depth 10, maximum depth 500, missing data per site < 20%, missing data per individual < 40%, minor allele frequencies per site > 10% and minimum count read for heterozygous genotype > 3. Sites with missing data on parents or segregation patterns in progenies that were not in agreement with the parental genotypes were also excluded. Dataset filtering and formatting were coded in R 3.4.4 (R core team 2017) and based on the vcfR 1.5.0 package (Knaus and Grünwald 2017) to import vcf files into the R environment.

Linkage analysis and parental map construction

SNPs with 1:1 Mendelian segregation if segregating only in one parent and 1:2:1 if segregating in both parents were retained for linkage analysis. SNPs with a significant segregation deviation within families (χ^2 test; $P < 0.001$) were eliminated. Finally, SNPs were thinned to maintain a minimum spacing of 100 bp between adjacent markers.

Linkage analysis and map constructions were conducted separately for each parent by population according to the cross-pollinated (CP) model in JoinMap 4.1 software (Van Ooijen 2012). In each family, the "hk × hk" segregation patterns were used for both parents, while "lm × ll" and "nn × np" segregation patterns were used for the female and male parental maps, respectively. Regarding the difference in dataset size, linkage groups were established using a grouping LOD threshold value of 7 for the parents of population A and 5 for those of population B. Parental maps were computed using recombination frequencies below 0.45, LODs over 1.0, a regression algorithm with two ordering rounds and the Kosambi mapping function.

QTL detection for sex determination

Sex was determined by looking at the type of inflorescence produced by each progeny. Phenotyping was conducted in 2016 and 2017 in both blocks to deal with the erratic flowering of *D. alata* and to avoid vine mixing issues. Overall, 69 progenies (32 females and 37 males) in population A (74F × Kabusa) and 75 progenies (31 females and 44 males) in population B (74F × 14 M) were phenotyped with confidence.

QTL detection was conducted on the four parental genetic maps using the R/qtl 1.42-8 package (Broman et al. 2003) and a simple interval mapping approach (options: step = 1 cM, error.prob = 1e−08, map.function = "kosambi", model = "binary", method = "hk"). Significance thresholds were calculated through permutations (1000) with an alpha risk of 0.05. QTL confidence intervals were computed using the "bayesint" function and 0.95 probability coverage of the interval.

Construction of the *D. alata* reference map and comparison with *D. rotundata*

As the two mapping populations were derived from crosses involving the same female (74F), an integrated map of the female parent (74F) was computed using the JoinMap "combine groups for map integration" function. Three genetic maps of female 74F were thus generated: from the population A dataset (74F_A), the population B dataset (74F_B) and the integrated map (74F). The final *D. alata* consensus map was constructed using this same function and starting from the integrated female linkage groups (74F) and both male parent groups (Kabusa and 14 M). Homology between the four parental maps, the integrated female map and the final consensus map were visualized using the R package ggplot2 2.1.1 (Wickham 2016).

The *D. alata* genome assembly v1 available as scaffolds accounting for roughly half of the genome and 80–90% of protein-coding loci (Water Yam Genome Project—ftp://yambase.org/genomes/Dioscorea_alata/) was anchored to our consensus map. To do that, SNP flanking sequences (60 bp upstream and 60 bp downstream around the variant position) were extracted using SNiPlay3 (Dereeper et al. 2015). These sequences were mapped on *D. alata* scaffolds using BLAST (Basic Local Alignment Search Tool, ncbi-blast v2.2.30).The results were parsed using an E-value threshold of 1e-10 and keeping secondary hits only if the difference [− log10(best hit E-value)] − [− log10(hit E-value)] was lower than 2.

The consensus map was also compared to the *D. rotundata* reference genome (pseudo-chromosomes BDMI0100001-21; Tamiru et al. 2017). The *D. rotundata* genome was divided into pieces cutting halfway between

SNPs included in our *D. alata* reference genetic map. The resulting genomic fragments were then reordered according to the SNP positions in the *D. alata* reference map developed in this study. Synteny between *D. alata* and *D. rotundata* was visualized using a Circos approach via the circlize R package 0.4.3 (Gu 2014).

Results

Genotyping-by-sequencing and SNP filtering

Overall, 121 progenies from population A (74F × Kabusa) and 193 from population B (74F × 14 M) were genotyped. Around 4.4 and 3.6 million reads per progeny were obtained with 83.9% and 82.5% of the reads mapping on the *D. rotundata* genome used as reference sequence, and for population A (74F × Kabusa) and population B (74F × 14 M), respectively (Online Resource 1). For the female (74F), 25 million reads were obtained, 84.6% of which were mapped. Nineteen million and 21 million reads were obtained, 82.2% and 84.2% of which were mapped for Kabusa and 14 M male parents, respectively. On average, 10% of the mapped reads were aligned to multiple positions (Online Resource 1) and removed.

By population, SNP filtering on the genotypic information quality (i.e., depth and allele frequencies) resulted in the detection of 29,224 and 11,808 SNPs in populations A and B, respectively (Table 1). By excluding sites based on missing data or discrepancies between segregation patterns in progenies and parental genotypes, 17,446 and 5434 SNPs were conserved for populations A and B, respectively. The segregation distortion threshold discarded 33% of SNPs in population A and 59% in population B (Table 1). Keeping a maximum of one SNP every 100 bp reduced the SNP dataset by approximately half. Thus, 5373 SNPs and 1075 SNPs were used for linkage analysis in population A and population B, respectively. Because of the missing data threshold per progeny, the final dataset included 79 progenies for population A and 110 progenies for population B.

Linkage analysis and parental maps

A total of 5837 unique SNPs were used in the linkage analysis, with 611 SNPs being common to both populations. Although a single female parent (74F) was used, common SNPs were not homogeneously distributed across segregation patterns between populations (Table 2). The proportions of markers segregating only in the female or male parent were similar in each population. Markers heterozygous in both parents (hk × hk) were more represented in population B (41%) than in population A (15%).

Four parental maps and an integrated female map were built. Because of the strong linkages revealed by the pairwise recombination frequencies and LOD scores (Online Resource 2), linkage groups were confidently defined for each parent. The number of linkage groups by parental map ranged from 21 for 74F_B to 26 for Kabusa. This was higher

Table 1 Summary of SNP filtering and dataset sizes per population

Population	High-quality SNPs	Low missing data and adequate segregation pattern	Undistorted SNPs	Dataset used in linkage mapping				
				No. of SNPs	No. of progenies	SNP depth[a]	NA per site[a] (%)	NA per progeny[a] (%)
A	29,224	17,446	11,667	5373	79	113	12.2	7.5
B	11,808	5434	2227	1075	110	88	14.2	8.6

The SNPs used in linkage mapping are the undistorted SNPs (χ^2 test; $P < 0.001$) thinned so that no two sites were within 100 base pairs

[a]Median

Table 2 Segregation type by mapping population

	Segregation type in population B				Total population A
	<hk × hk>	<lm × ll>	<nn × np>	Absent	
Segregation type in population A					
<hk × hk>	92	57		683	832 (15%)
<lm × ll>	188	194		1814	2196 (41%)
<nn × np>			80	2265	2345 (44%)
Absent	162	93	209		
Total population B	442 (41%)	344 (32%)	289 (27%)		

Segregation types are in JoinMap format. In brackets, percentage of the total number of SNP by population

than the *D. alata* base chromosome number (i.e., 20) and may have been the result of the separation of linkage groups containing only a few markers. Nevertheless, the integrated female map (74F) contained 20 linkage groups built using information on both population datasets for each integrated linkage group.

Maps of population B parents were smaller and less dense than those of population A parents. Parental map lengths ranged from 1227 cM for 14 M to 2348 cM for 74F_A, respectively. The map density ranged from one SNP every 2.1 cM in the Kabusa map to one SNP every 3.7 cM in the 74F_B map (Table 3).

Map integration and consensus map construction

Male linkage groups were combined with the integrated female map based on the recombination frequencies (Online Resources 3–4). Linkage groups were numbered in reference to the *D. rotundata* genome, as for the integrated female map (74F). Map integration was fairly accurate, as revealed by the good collinearity between homolog linkage groups from the different parental maps (Online Resource 5).

The consensus genetic map obtained in this study spanned 2613.5 cM and contained 1579 SNPs distributed on 20 linkage groups, as expected (Fig. 1, Table 4). Linkage groups contained from 20 (LG14) to 145 (LG05) SNPs, with a genetic length ranging from 55.6 cM (LG14) to 188.1 cM (LG05). The mean marker density was one SNP every 1.68 cM. Each SNP had a single position (Table 4).

One linkage group of the 74F_A map that could not be included in the integrated female map but had sufficient common SNPs with male maps was also integrated in the final consensus map (Online Resource 3). The consensus map did not contain information of male B (14 M) for three linkage groups (LG3, LG4 and LG16) as no homologs were found (Online Resource 4). The male A (Kabusa) homolog linkage group of LG14 was constructed but did not contain sufficient bridge markers with the other three parental homologs of LG14 to be integrated into the final consensus map (Online Resource 4).

For LG6, two maps were included in the final consensus map: the integrated female map of LG6 (LG6_F) and a male consensus map of LG6 (LG6_M). Indeed, map integration of the different LG6 homologs was not possible because the male maps did not contain any common SNPs with the female consensus map. As dioecy can be related to the chromosome architecture, the relation between these map integration issues and sex determination was further investigated.

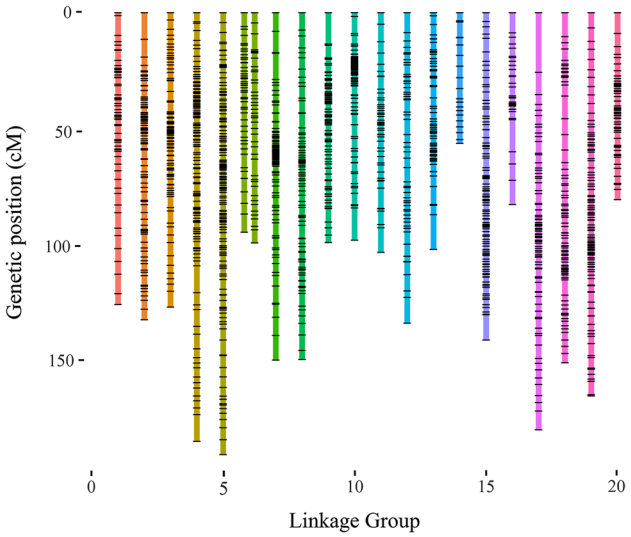


Fig. 1 Greater yam (*D. alata* L.) consensus genetic map containing 1548 SNPs. X-axis, linkage groups numbered from LG1 to LG20 homology with the *D. rotundata* reference genome (Tamiru et al. 2017); y-axis, genetic distance (Kosambi mapping function; cM)

Detection of sex-determining QTLs

The phenotypic data analysis revealed no significant differences between the observed sex ratio and a 1:1 theoretical ratio within both populations (χ^2 test; population A, P value = 0.55; population B, P value = 0.13), indicating that sex determination may be controlled by one dominant allele.

Combining genotypic and phenotypic datasets, QTL detection was performed on fewer progenies and conducted on parental maps with 42 (19 females + 23 males) and 60 (27 females and 33 males) progenies for populations A and B, respectively. Only one QTL per population was detected on

Table 3 Summary of parental maps

Parent	Map	No. of LGs[1]	No. of SNPs	Length (cM)	Average marker interval (cM)
Female	74F_A	25 (21)	1035	2348	2.30
	74F_B	21 (20)	486	1705	3.70
	74F	20	983	2120	2.20
Male	Kabusa	26 (21)	1078	2195	2.14
	14M	21 (17)	371	1227	3.50

In brackets, the number of linkage groups integrated into the final consensus map

Table 4 Description of the consensus genetic map for *D. alata* L. by linkage groups

Linkage group	No. of SNPs	Genetic length (cM)	Marker interval (cM)[a]	Max gap (cM)	No. of gaps > 5 cM	Physical length (Mb)[b]	Marker interval (kb)[c]
LG01	56	124.2	2.26	8.5	6	28.0	142.3
LG02	88	130.6	1.50	11.3	2	33.9	141.7
LG03	81	125.3	1.57	6.7	3	18.6	120.0
LG04	125	182.5	1.47	11.4	5	29.9	126.7
LG05	145	188.1	1.31	6.4	4	28.5	99.8
LG06_M	47	93.5	2.03	7.7	4	20.8	370.3
LG06_F	42	98.1	2.39	7.0	3	19.4	173.2
LG07	88	147.8	1.70	11.2	8	19.3	123.0
LG08	80	147.6	1.87	7.4	8	20.3	117.0
LG09	90	98.0	1.10	13.3	3	27.6	133.3
LG10	72	96.9	1.36	13.7	3	18.1	88.6
LG11	47	102.1	2.22	10.5	5	15.7	209.1
LG12	56	132.1	2.40	11.0	9	28.9	73.5
LG13	65	100.9	1.58	14.8	2	24.6	220.6
LG14	20	55.6	2.93	6.6	1	11.7	139.7
LG15	102	139.3	1.38	10.8	2	18.5	99.6
LG16	31	81.7	2.72	13.7	5	2.8	26.4
LG17	83	177.5	2.17	25.4	4	21.3	165.1
LG18	88	149.0	1.71	10.2	6	29.4	133.2
LG19	117	163.2	1.41	6.5	4	39.7	132.4
LG20	56	79.6	1.45	8.8	2	13.0	89.1
Total	1579	2613.5	1.68			442.1	139.1

For the sex-related LG6, two maps were conserved: the female-integrated map (LG6_F) and a male consensus map (LG6_M)

[a]Mean distance between SNPs

[b]Total length after reordering the *D. rotundata* reference genome (Tamiru et al. 2017) according to *D. alata* consensus map

[c]Median physical distance between SNPs in the reordered *D. rotundata* genomic sequence

LG6 homologs and only in the male maps (Online Resource 6A). No QTLs were detected using the female maps. In agreement with the sex ratio observed within both populations, these findings suggest that only one locus may be involved in sex determination and may be inherited via a system of XX/XY sex chromosomes involving heterogametic males (XY).

For male A (Kabusa), the QTL confidence interval spanned from 1.1 to 30.2 cM, with a peak LOD score of 5.2 located at 13.0 cM. The nearest SNP (06.1_27885348) located at 13.3 cM allowed us to predict sex in 85% of the cases (Fig. 2). For male B (14 M), the QTL confidence interval spanned from 0 to 34.8 cM, with a peak LOD score of 3.8 located at 1.0 cM. The LOD drop in the middle of the QTL interval was due to a marker phase change. The nearest SNP (06.1_27950405) located at 0.0 cM allowed us to predict sex in 77% of the cases (Fig. 2). In both maps, the QTL peaks contained tightly linked SNPs (Online Resource 6B). In both cases, no secondary peaks were found when the most significant SNPs were used as covariates. Once projected on the consensus male map (LG6_M), the two QTL intervals co-localized and were located between 4.8 and 33.9 cM and between 0 and 34.8 cM for male A and male B, respectively. The locus involved in sex determination in the two populations may be the same.

D. alata scaffold anchoring and synteny with *D. rotundata*

Based on the SNP positions in our consensus *D. alata* genetic map compared to SNP positions in the *D. rotundata* genome, the *D. rotundata* genome was reordered: (i) to highlight possible chromosome rearrangements between the two species and (ii) to estimate the relationship between physical and linkage map distances in *D. alata*.

The mean homology between *D. alata* and *D. rotundata* linkage groups was 87% when computed as the percentage of SNPs located in a *D. alata* linkage group and in its *D. rotundata* homolog (Online Resource 7). The highest homology between the two yam species was found for LG5

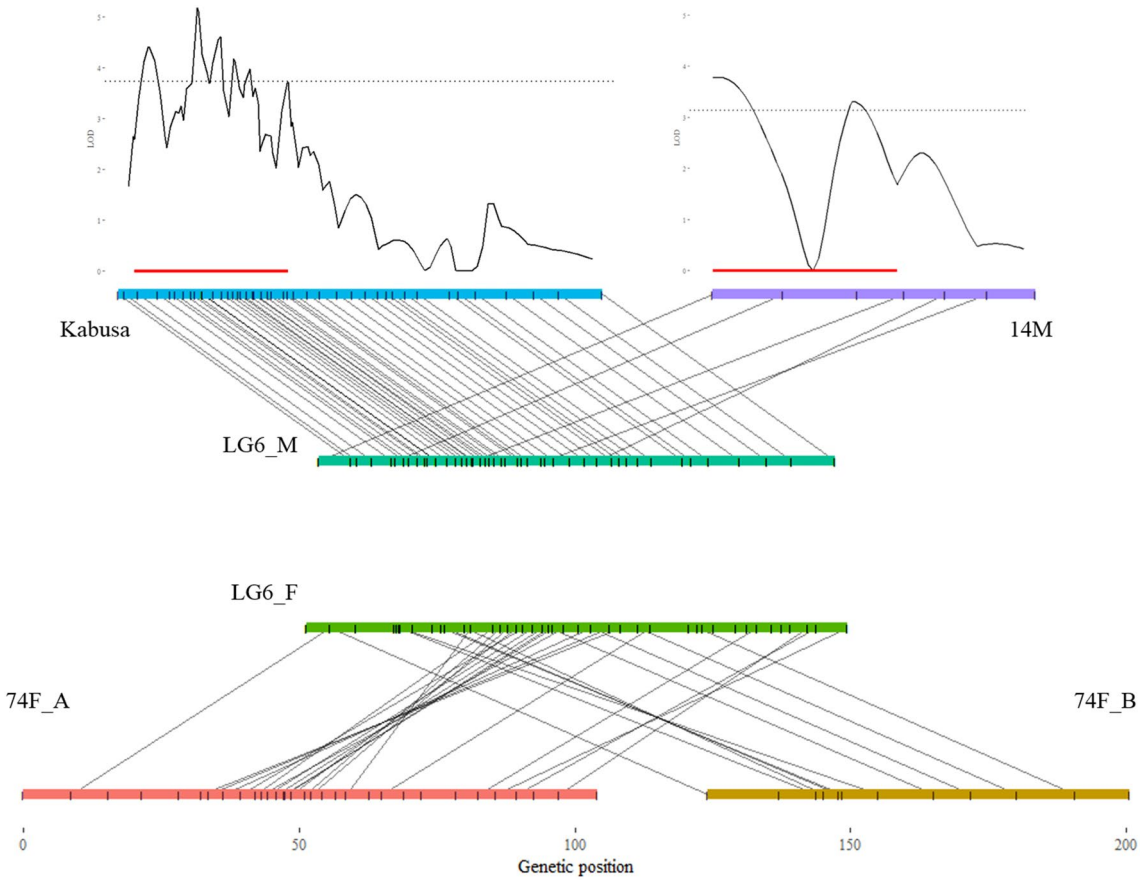


Fig. 2 Homology between parental genetic maps of LG6 homologs and detection of the sex-determining QTLs. On LOD score plots: dashed lines, LOD score threshold; red lines, QTL confidence interval. On linkage groups: black ticks, SNPs position

and LG8, with 99% of SNPs from *D. alata* belonging to the same respective linkage group in *D. rotundata* (Fig. 3; Online Resource 7).

D. alata LG19 was the less conserved, with only 48% of SNPs belonging to *D. rotundata* chromosome 19 (Online Resource 7). This could partially be explained by the fact that during *D. rotundata* sequencing a supernumerary chromosome 21 was assumed which was mapped at the end of our LG19 (Fig. 2). In our study, LG19 also included 10% of SNPs located on chromosome 16 of *D. rotundata* (Online Resource 6) while spreading over 63% of this chromosome (Online Resource 8). In our results, LG16 contained few SNPs (31) and was the smallest, with a total estimated length of 2.8 Mb (Table 4; Fig. 3).

LG15 and LG7 also appeared to be rearranged regarding the percentage of SNPs (Online Resource 7). This was even truer for LG15 when homology was estimated on the basis of the percentage of genomic sequence mapped on its homolog chromosome (Online Resource 8). Indeed, the reordered chromosome corresponding to LG15 was composed of *D. rotundata* chromosomes 15 (43%), 6 (26%) and 4 (11%) (Fig. 3).

The median SNP interval was 139.1 kb in our reordered *D. rodundata* genome (Table 4). SNP intervals ranged from 26.4 kb for LG16 to 370.3 kb for LG6_M. For most linkage groups, the relationship between the physical and the linkage map distances varied along linkage groups, in agreement with an expected chromosomal structure made of highly recombinant telomeres and a less recombinant centromere (Online Resource 9).

D. alata has not yet been completely sequenced. However, useful genomic resources have been released upon which our SNPs could be positioned. Overall, 743 scaffolds from the *D. alata* genome v1 were anchored in the consensus maps, representing around 40% (115.8 Mb/287.3 Mb) of the total scaffold length (Online Resource 10). Around 38.5% (286/743) of these scaffolds could theoretically be oriented as they contained two or more SNPs. The scaffolding and linkage analysis findings seemed quite consistent, as revealed by the fact that 74.5%

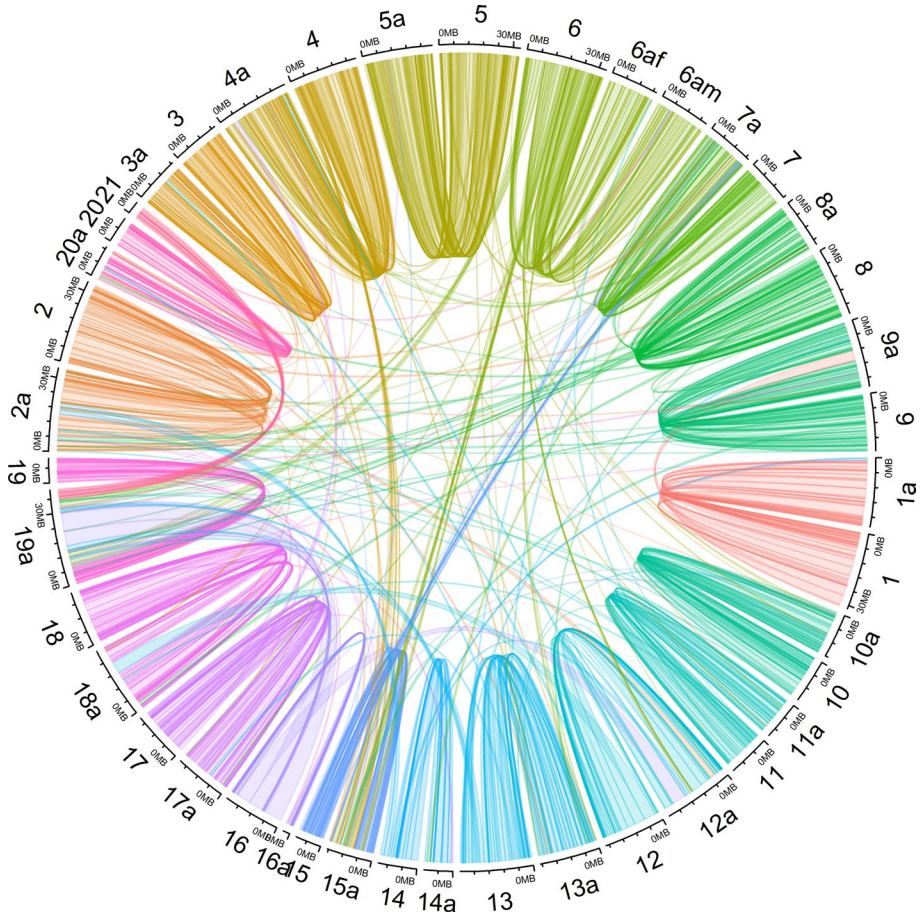


Fig. 3 Synteny between *D. alata* and *D. rotundata*. Chromosomes numbered with the suffix 'a' are for *D. alata* corresponding to the *D. rotundata* reference sequence reordered according to our final consensus map. For LG6, both female (6af) and male (6am) reordered chromosomes were conserved

(213/286) of the scaffolds containing two or more SNPs contained only SNPs mapped on the same LG.

Discussion

A *D. alata* reference map

The method used to build the reference genetic map was designed to be conservative. Indeed, integration of maps into the final consensus map was based on pairwise recombination frequencies in parental datasets and not on map projections. All SNPs mapped in parental maps were thus not necessarily included in the final reference map and vice versa. Although this procedure is more restrictive, it was applied to minimize the impact of errors due to the reversal of locations for short distance markers and/or of structural variations (Mace et al. 2009; Khan et al. 2012). In the same conservative spirit, highly distorted markers were eliminated before linkage analysis (Zhang et al. 2010).

Our consensus map is the first SNP high-density genetic map for *D. alata*. It contains 20 linkage groups, as expected, and 1579 SNPs spread over 2613.5 cM, with an average marker interval of 1.7 cM. Indeed, using the same Kosambi mapping function, the previously published map lengths were 1233 cM, 1538 cM and 3229.5 cM, and they contained 494, 523 and 380 markers with an average marker spacing of 2.6 cM, 3.3 cM and 14.2 cM according to Mignouna et al. (2002), Petro et al. (2011) and Bhattacharjee et al. (2018), respectively. The genome coverage was estimated at 65% by Mignouna et al. (2002) and 80% by Petro et al. (2011). The estimated genome coverage of our map reached 94% using Method 4 of Chakravarti et al. (1991).

Diploid *D. alata* genotypes have a haploid genome size estimated to be between 1C = 0.46 pg = 454 Mb (Arnau et al. 2009) and 1C = 0.57 pg = 562 Mb (Obidiegwu et al. 2010) by flow cytometry analysis. Saski et al. (2015) de novo sequencing resulted in an assembly of contigs covering 428.9 Mb. These estimations are slightly smaller than the *D. rotundata* genome size (570–579 Mb; Tamiru et al. 2017). Moreover, physical distance in *D. rotundata* and *D. alata* seems well correlated and proportional to smaller distance for *D. alata* ($r^2 = 0.79$; coefficient of proportionality = 0.786; Online Resource 11). The good resolution (in kb) of our map thus may have been fairly well estimated or even slightly underestimated in this study.

In this sense, our estimated marker density of one SNP every 139.1 kb agreed with the fact that around 40% of the

total assembly of the *D. alata* genome v1 was contained in our reference map. Indeed, presently 50% of this pre-release assembly is composed of scaffolds longer than 145.7 kb (Water Yam Genome Project—ftp://yambase.org/genomes/Dioscorea_alata/). In *D. rotundata*, 89.6% of the genome assembly is included in scaffolds longer than 200 kb (Table S3 in Tamiru et al. 2017). If the ongoing *D. alata* sequencing results in similar scaffolding, our reference map may be sufficient to order most of the future assembly.

More generally, this genetic map for greater yam, associated with the reordered *D. rotundata* genome and the mapping of available *D. alata* scaffolds, will promote further investigations on the inheritance of key traits and the development of molecular breeding tools. It will also help gain further insight into yam evolution and facilitate the transfer of knowledge regarding different yam species. We thus strongly encourage retaining the linkage group nomenclature we used for *D. alata* as it is also based on *D. rotundata* nomenclature.

Segregation distortion

The difference in linkage analysis power between population A and population B mostly resulted from the lower number of SNPs available for mapping in population B. Three consensus linkage groups of the final consensus map did not contain information on the population B male parent (14 M) as no homologs were found. Moreover, parental maps derived from population B were shorter and less dense than those derived from population A.

First, for a similar GBS quality (i.e., number of produced reads and missing data), fewer SNPs were detected in population B than in population A. This agreed with the genetic proximity between 74F and 14 M (76% of shared alleles) compared to that of 74F and Kabusa (65% of shared alleles). Then, the segregation distortion threshold excluded a higher proportion of markers in population B than in population A. Segregation distortions have been widely reported in plant species including *D. alata*. In Petro et al. (2011), 19% of markers were tagged as distorted, while in Bhattacharjee et al. (2018) it was 39.8%. However, comparisons are limited due to the unknown threshold used to test distortion.

Regarding the decrease in the number of progenies from hybridization to mapping datasets, one major hypothesis could be proposed. Indeed, the main differences were the ratio of introduced embryos to the number of pollination between population A (140/250 = 56%) and population B (280/360 = 78%), which was related to gametophyte selection, and the proportion of rescued embryos successfully brought to the field (population A, 121/140 = 86%; population B, 193/280 = 69%), which was related to early-stage zygotic selection. However, the sampling bias—estimated when considering the proportion of non-genotyped progenies and introduced by filters on data quality per progeny—was similar between the two populations. Thus, we hypothesized that early-stage zygotic selection affected to a greater extent population B for which segregation distortion was mainly related to genes involved in the response to in vitro and field stresses. Regarding the smaller size of population A compared to population B, segregation distortion may have been more related to the sampling bias within population A. This hypothesis agrees with the relative genetic proximity of 74F and 14 M, and the inbreeding depression observed for seed germination and zygotic viability in *D. alata* breeding programs (Abraham et al. 2006).

Sex determination in Dioscorea

Dioecy is a key character in *Dioscorea* species (Fig. 2 in Viruel et al. 2016). Based on cytological observations, previous studies have mostly reported an XX/XY chromosome system (review in Martin 1966), indicating that Y is the sex-determining chromosome and males are heterogametic. When assessing the sex ratio in a test-cross design, an XX/XY chromosome system was also proposed for *D. floribunda* (Martin 1966) and for the dioecious *D. tokoro* using QTL detection with AFLP genetic maps (Terauchi and Kahl 1999). Our results agreed with a XX/XY sex-determining system in *D. alata* mapped at the same location (beginning of LG6_M) in the two male parents. This conclusion has to be confirmed using a more diverse range of genotypes to ensure that the XX/XY system we discovered was not specific to the female parent (74F).

The main exception to the XX/XY system was found for the trioecious (mostly dioecious) *D. rotundata* species (Tamiru et al. 2017) for which a ZW/ZZ system (heterogametic female) mapped at the beginning of pseudo-chromosome 11 was described. Indeed, the authors conducted bulk segregant analysis in a biparental population and identified SNPs linked to sex heterozygous in the female parent but not in the male parent (see Fig. 4C in Tamiru et al. 2017). They also identified female-specific regions for which they developed a PCR primer pair. Interestingly, the beginning of pseudo-chromosome 11, which is linked to sex determination in *D. rotundata*, seemed to be rearranged in *D. alata* as it mapped to LG11, but also to LG2, LG18, LG19 and LG6_M (Online Resource 10). The change from cosexuality to dioecy implies a complex evolutionary process consisting of successive mutations for male and female sterility and sex chromosome rearrangement (Charlesworth 2002, 2015; Otto et al. 2011; Kumar et al. 2014). Moreover, the mostly dioecious species *D. rotundata* belongs to the same section (Enantiophyllum) as the strictly dioecious *D. alata*, and there is good synteny and sequence homology between the two species. We thus suggest that their sex determination

systems may be related (e.g., transition from XY into ZW system; Kumar et al. 2014).

So far, heteromorphic sex chromosomes have been identified in around half of the species for which sex chromosomes were detected (Hobza et al. 2017). Heteromorphic sex chromosomes have been reported in *Dioscorea* (Martin 1966), but there is still no cytological evidence due to the small size of *Dioscorea* chromosomes. Our results revealed a sex-linked QTL interval larger than 30 cM on the Y chromosome (LG6_M) spreading over approximately 10 Mb. The size of the confidence interval could be related to the small population size used for QTL detection due to the erratic flowering pattern of *D. alata* (Malapa et al. 2005). However, it could also be related to a sex-linked region with a low recombination rate. Indeed, the establishment of sex-determining regions associated with local suppression of recombination in Y chromosomes is the key driver of chromosome Y differentiation (Otto et al. 2011; Kumar et al. 2014; Hobza et al. 2015). In this sense, LG6_M (Y) is the linkage group with the lowest resolution (kb^{-1}) and contains no common marker with LG6_F (X). In *D. tokoro*, Terauchi and Kahl (1999) showed that all markers of the Y chromosome spreading over 23.3 cM showed tight linkage to sex compared to the absence of sex-linked markers on the X chromosome spreading over 82.5 cM.

Conclusion

Linkage analysis studies on two biparental populations were combined to build a high-density SNP genetic map of greater yam (*D. alata*). This map covered 94% of the genome and contained 1579 SNPs. Regarding sequence homology and synteny with its already sequenced relative *D. rotundata*, a reordered *D. rotundata* genome adapted to *D. alata* was proposed. The goal was: (i) to facilitate further investigations on the identification of loci linked to key traits, history and evolution and (ii) to enhance knowledge transfer within the *Dioscorea* genus. Indeed, the estimated resolution of this map was 139.1 kb, thus allowing QTL and gene cloning strategies. Based on our study, we also encourage a common linkage group nomenclature. Information on female and male LG6 carrying a major locus determining sex was separately conserved within this consensus map. Indeed, sex ratio analysis within populations and QTL detection revealed a XX/XY sex chromosome system, and the presence of heteromorphic sex chromosomes could reasonably be hypothesized.

Author contribution statement GA designed the study with the support of HC. EM, EN, GA and MCG created and maintained the plant material. FC, HC, HV and PM generated the genotyping-by-sequencing data and their analysis. EN, EM and FC phenotyped the progenies. FC and GA conducted linkage analysis. FL and FC performed the QTL mapping. FC, GA, HC, PM and FL wrote the manuscript.

Acknowledgments This work was financially supported by the AfricaYam project (Grant OPP1052998—Bill and Melinda Gates Foundation) and the European Union and Guadeloupe Region (Programme Opérationnel FEDER—Guadeloupe—Conseil Régional 2014–2017). Floriane Lawac received a grant from the French Embassy in Vanuatu and the Vanuatu Agricultural Research and Technical Center (VARTC) for her MSc research. The authors would like to thank Ronan Rivallan, Louis Minfort and Guillaume Martin for their help on genotyping, phenotyping and GBS analysis, respectively. A special thanks to Brigitte Courtois for her helpful discussions and manuscript improvement and to David Manley for English proofing.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards The authors declare that the experiments comply with the current laws of the country in which they were performed.

Data accessibility statement Datasets generated in the current study are posted at Online Resource 10 (xlsx file). The Illumina HiSeq 3000 sequencing raw data are available in the NCBI SRA (Sequence Read Archive), under the BioProject number: PRJNA515897. KeyGene N.V. owns patents and patent applications protecting its sequence-based genotyping technologies.

References

Abang MM, Hoffmann P, Winter S, Green KR, Wolf GA (2004) Vegetative compatibility among isolates of Colletotrichum gloeosporioides from yam (*Dioscorea* spp.) in Nigeria. J Phytopathol 152:21–27

Abraham K, Gopinathan Nair P (1990) Floral biology and artificial pollination in *Dioscorea alata* L. Euphytica 48:45–51

Abraham K, Sreekumari MT, Sheela MN (2006) Seed production strategies and progeny selection in greater Yam breeding. In: 14th Triennial symposium of the international society for tropical root crops, India

Andris M, Aradottir G, Arnau G et al (2010) Permanent genetic resources added to molecular ecology resources database 1 June 2010–31 July 2010. Mol Ecol Resour 10:1106–1108

Arnau G, Nemorin A, Maledon E, Abraham K (2009) Revision of ploidy status of *Dioscorea alata* L. (Dioscoreaceae) by

cytogenetic and microsatellite segregation analysis. Theor Appl Genet 118:1239–1249
Arnau G, Nemorin A, Maledon E, Nudol E (2011) Advances on polyploid breeding in yam *D. alata*. In: Proceeding of the first international symposium on roots, Rhizomes, Tubers, Plantains, Bananas and Papaya. 7–10 November 2011 Santa Clara, Cuba
Arnau G, Bhattacharjee R, Sheela MN, Chair H, Malapa R, Lebot V et al (2017) Understanding the genetic diversity and population structure of yam (*Dioscorea alata* L.) using microsatellite markers. PLoS ONE 12(3):e0174150
Ayensu ES, Coursey DG (1972) Guinea yams the botany, ethnobotany, use and possible future of yams in West Africa. Econ Bot 26:301–318
Bhattacharjee R, Nwadili CO, Saski CA et al (2018) An EST-SSR based genetic linkage map and identification of QTLs for anthracnose disease resistance in water yam (*Dioscorea alata* L.). PLoS ONE 13(10):e0197717
Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890
Chakravarti A, Lasher LK, Reefer JE (1991) A maximum likelihood method for estimating genome length using genetic linkage data. Genetics 128(1):175–182
Charlesworth D (2002) Plant sex determination and sex chromosomes. Heredity 88(2):94–101
Charlesworth D (2015) Plant contribution to our understanding of sex chromosome evolution. New Phytol 208:52–65
Coursey DG (1967) Yams. Longmans, London, p 230
Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510
DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498
Dereeper A, Nicolas S, Lecunff L et al (2015) SNiPlay: a web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects. BMC Bioinformatics 12:134
Egesi CN, Asiedu R (2002) Analysis of yam yields using the additive main effects and multiplicative interaction (AMMI) model. Afr Crop Sci J 10:195–201
Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379
Garsmeur O, Droc G, Antonise R et al (2018) A mosaic monoploid reference sequence for the highly complex genome of sugarcane. Nat Commun 9:2638
Gu Z (2014) Circlize implements and enhances circular visualization in R. Bioinformatics 30:2811–2812
Herten K, Hestand MS, Vermeesch JR, Van Houdt J (2015) GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. BMC Bioinformatics 16:73
Hobza R, Kubat Z, Cegan R, Jesionek W, Vsykot B, Kejnovsky E (2015) Impact of repetitive DNA on sex chromosome evolution in plants. Chromosome Res 23:561–570
Hobza R, Cegan R, Jesionek W, Vyskot B, Kubat Z (2017) Impact of repetitive elements on the Y chromosome formation in plants. Genes 8:302
Hochu I, Santoni S, Bousalem M (2006) Isolation, characterization and cross-species amplification of microsatellite DNA loci in the tropical American yam *Dioscorea trifida*. Mol Ecol Notes 6:137–140
Khan MA, Han Y, Zhao YF, Troggio M, Korban SS (2012) A multi-population consensus genetic map reveals inconsistent marker order among maps likely attributed to structural variations in the apple genome. PLoS ONE 7(11):e47864
Knaus BJ, Grünwald NJ (2017) VCFR: a package to manipulate and visualize variant call format data in R. Mol Ecol Resour 17:44–53
Kumar S, Kumari R, Sharma V (2014) Genetics of dioecy and causal sex chromosomes in plants. J Genet 93:241–277
Mace ES, Rami JF, Bouchet S et al (2009) A consensus genetic map of sorghum that integrates multiple component maps and high-throughput Diversity Array Technology (DArT) markers. BMC Plant Biol 9:13
Martin FW (1966) Sex ratio and sex determination in Dioscorea. J Hered 57:95–99
Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal 17:10–12
Mignouna HD, Mank RA, Ellis THN et al (2002) A genetic linkage map of water yam (*Dioscorea alata* L.) based on AFLP markers and QTL analysis for anthracnose resistance. Theo Appl Genet 105:726–735
Misuki I, Tani N, Ishida K, Tsumura Y (2005) Development and characterization of microsatellite markers in a clonal plant, *Dioscorea japonica* Thunb. Mol Ecol Notes 5:721–723
Muzac-Tucker I, Asemota HN, Ahmad MH (1993) Biochemical composition and storage of Jamaican yams (*Dioscorea* spp.). J Sci Food Agric 62(3):219–224
Obidiegwu J, Rodriguez E, Ene-Obong E et al (2010) Ploidy levels of *Dioscorea alata* L. germplasm determined by flow cytometry. Genet Resour Crop Evol 57(3):351–356
Otto SP, Pannell JR, Peichel CL et al (2011) About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. Trends Genet 27:358–367
Petro D, Onyeka TJ, Etienne S, Rubens S (2011) An intraspecific genetic map of water yam (*Dioscorea alata* L.) based on AFLP markers and QTL analysis for anthracnose resistance. Euphytica 179:405–416
Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE. https://doi.org/10.1371/journal.pone.0032253
R Core Team (2017) R: a language and environment for statistical computing. https://www.R-project.org/
Risterucci AM, Hippolyte I, Perrier X et al (2009) Development and assessment of diversity arrays technology for highthroughput DNA analyses in Musa. Theor Appl Genet 119:1093–1103
Sartie A, Asiedu R (2014) Segregation of vegetative and reproductive traits associated with tuber yield and quality in water yam (*Dioscorea alata* L.). Afric J Biotech 13(28):2807–2818
Saski CA, Bhattacharjee R, Scheffler BE, Asiedu R (2015) Genomic resources for water yam (*Dioscorea alata* L.): analyses of EST-sequences, de novo sequencing and GBS libraries. PLoS ONE 10(7):e0134031
Tamiru M, Natsume S, Takagi H et al (2017) Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular marker for sex determination. BMC Biol 15:86
Terauchi R, Kahl Günter (1999) Mapping of the *Dioscorea tokoro* genome: AFLP markers linked to sex. Genome 42:757–762
Terauchi R, Konuma A (1994) Microsatellite polymorphism in Dioscorea tokoro, a wild yam species. Genome 37:794–801
Tostain S, Scarcelli N, Brottier P, Marchand JL, Pham JL, Noyer JL (2006) Development of DNA microsatellite markers in tropical yam (*Dioscorea* sp.). Mol Ecol Notes 6:173–175
Van Ooijen JW (2012) JoinMap 4.1, software for the calculation of genetic linkage maps in experimental populations of diploid species. Kyazma B.V., Wageningen
Viruel J, Segarra-Moragues JG, Raz L et al (2016) Late cretaceous-early eocene origin of yams (Dioscorea, Dioscoreaceae) in the Laurasian Palaearctic and their subsequent Oligocene-Miocene diversification. J Biogeogr 43:750–762
Ward JA, Bhangoo J, Fernández-Fernández F et al (2013) Saturated linkage map construction in Rubus idaeus using genotyping by

sequencing and genome-independent imputation. BMC Genom 14:2
Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer, New York
Wilkin P, Schols P, Chase MW et al (2005) A plastid gene phylogeny of the yam genus, *Dioscorea* roots, fruits and madagascar. Syst Bot 30:736–749
Zhang L, Wang S, Li H et al (2010) Effects of missing marker and segregation distortion on QTL mapping in F2 populations. Theor Appl Genet 121:1071

Publication 4

• Ehounou, A. E., Cornet, D., Desfontaines, L., Marie-Magdeleine, C., Maledon, E., Nudol, E., Beurier, G., Rouan, L., Brat, P., Lechaudel, M., Nous, C., N'Guetta, A. S. P., Kouakou, A. M., & Arnau, G. (2021). Predicting quality, texture and chemical content of yam (*Dioscorea alata* L.) tubers using near-infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 29(3), 128-139.

Original research article

JNIRS

Journal of Near Infrared Spectroscopy
2021, Vol. 29(3) 128–139

DOI: 10.1177/09670335211007575
journals.sagepub.com/home/jns
SAGE

Predicting quality, texture and chemical content of yam (*Dioscorea alata* L.) tubers using near infrared spectroscopy

Adou Emmanuel Ehounou[1,2], Denis Cornet[3,4], Lucienne Desfontaines[5], Carine Marie-Magdeleine[6], Erick Maledon[4,7], Elie Nudol[4,7], Gregory Beurier[3,4], Lauriane Rouan[3,4], Pierre Brat[4,8], Mathieu Lechaudel[4,8], Camille Nous[9], Assanvo Simon Pierre N'Guetta[1,2], Amani Michel Kouakou[2] and Gemma Arnau[4,7]

Abstract

Despite the importance of yam (*Dioscorea* spp.) tuber quality traits, and more precisely texture attributes, high-throughput screening methods for varietal selection are still lacking. This study sets out to define the profile of good quality pounded yam and provide screening tools based on predictive models using near infrared reflectance spectroscopy. Seventy-four out of 216 studied samples proved to be moldable, i.e. suitable for pounded yam. While samples with low dry matter (<25%), high sugar (>4%) and high protein (>6%) contents, low hardness (<5 N), high springiness (>0.5) and high cohesiveness (>0.5) grouped mostly non-moldable genotypes, the opposite was not true. This outline definition of a desirable chemotype may allow breeders to choose screening thresholds to support their choice. Moreover, traditional near infrared reflectance spectroscopy quantitative prediction models provided good prediction for chemical aspects ($R^2 > 0.85$ for dry matter, starch, protein and sugar content), but not for texture attributes ($R^2 < 0.58$). Conversely, convolutional neural network classification models enabled good qualitative prediction for all texture parameters but hardness (i.e. an accuracy of 80, 95, 100 and 55%, respectively, for moldability, cohesiveness, springiness and hardness). This study demonstrated the usefulness of near infrared reflectance spectroscopy as a high-throughput way of phenotyping pounded yam quality. Altogether, these results allow for an efficient screening toolbox for quality traits in yams.

Keywords

Yam (*Dioscorea alata* L.), quality, texture, near infrared spectrometry, convolutional neural network

Received 27 January 2021; accepted 7 March 2021

Introduction

Yams (*Dioscorea* spp.), which are important crops for food security, are grown in tropical and sub-tropical regions.[1] They are dioecious herbaceous vines belonging to a C3 monocotyledonous genus cultivated for their starchy tubers.[2] *D. rotundata* and *D. alata* are the most important cultivated species and both belong to the same botanical section, i.e. Enantiophyllum. Greater yam (*D. alata*) ranks second in production importance, but it is the most widely cropped species in the world.[3] This species offers unique advantages in terms of potential yield, especially under low-fertility soil conditions (e.g. ease of propagation, early vigor for weed suppression and storability of tubers).[4,5] Yams are consumed in several forms (e.g. boiled, pounded, fried, baked, and roasted). In West Africa, where over 95% of the

[1]Université Félix Houphouët Boigny, UFR Biosciences, Abidjan, Côte d'Ivoire
[2]CNRA, Station de Recherche sur les Cultures Vivrières, Bouaké, Côte d'Ivoire
[3]CIRAD, UMR AGAP Institut, F-34398 Montpellier, France
[4]UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, F-34398 Montpellier, France
[5]INRAE, UR 1321 ASTRO Agrosystèmes tropicaux. Centre de recherche Antilles-Guyane, Petit-Bourg, France
[6]INRAE, UR 0143 URZ Unité de Recherches Zootechniques. Centre de recherche Antilles-Guyane, Petit-Bourg, France
[7]CIRAD, UMR AGAP Institut, Petit-Bourg, Guadeloupe, France
[8]CIRAD, UMR Qualisud, Capesterre-Belle-Eau, Guadeloupe, France
[9]Cogitamus Laboratory, Laboratoire Cogitamus, Montpellier, France

Corresponding author:
Denis Cornet, CIRAD Centre de Montpellier, UMR Agap Institut, F-34398 Montpellier, France.
Email: denis.cornet@cirad.fr

world's yams are produced,[6] pounded yam is the preferred consumption form. It is prepared by pounding the cooked tuber and kneading it into a sticky paste.[7] In this respect, *D. rotundata* is usually preferred to *D. alata* owing to its ease of dough formation when pounded.[7] However, several studies have shown that some genotypes of *D. alata* have the ability to form good dough, comparable to that of *D. rotundata*.[8,9]

The importance of yams in terms of food security has led to the establishment of several genetic improvement programs for *D. alata* throughout the tropical regions, in order to develop new varieties with high yield, good food quality and resistance to pests and diseases.[3,10–13] Sensory quality evaluations of new genotypes are mostly carried out using sensory panels. It is a laborious and lengthy process that lacks a high-throughput phenotyping method. For this reason, the organoleptic qualities of new varieties are assessed at the end of the breeding process. By that time, only a few genotypes remain and the chance of identifying high quality varieties is low. As a result, improved varieties are rarely adopted by end-users.

Organoleptic properties depend on several physico-chemical and textural characteristics. Dry matter, starch, proteins and sugar contents are important quality traits. Martin (1974) observed in Puerto Rico (USA) that high dry weights are associated with fine structure, dense feel and high quality.[14] Lebot et al (2005) observed that varieties with good boiled quality are characterized by high dry matter, starch and amylose contents.[15] Physico-chemical analysis studies showed that varieties with good poundability are also characterized by high dry matter contents.[16,17] Lastly, Brunnschweiler (2004) showed that pounded *D. rotundata* exhibited greater firmness, elastic recovery, gumminess and cohesiveness than *D. alata* varieties.[18] Several sensorial tests using trained or hedonic testing have been carried out in different West African countries in order to identify which texture and taste attributes of pounded yam are the most important. Five parameters are often mentioned in relation to quality, namely firmness, ease of molding, smooth appearance (absence of lumps), springiness and sweetness.[8,16,19–21]

The textural properties of pounded yam can be measured using a double compression test with a Textural Profile Analyser.[22] This method can be used to quantify different parameters (e.g. hardness, cohesiveness and springiness) linked to texture attributes. It has been successfully used to determine the textural quality of various food products, such as soybean-derived gels, banana, cassava roots or yam tubers.[23–27] However, chemical and textural characterization (both sensorial and TPA analysis) are costly, laborious and time-consuming. A high-throughput phenotyping method is required to be able to screen large numbers of breeding lines. In this regard, near infrared reflectance (NIR) spectroscopy has become a widely used method of quality control in the food processing industry.[28,29] It is a rapid, cost-effective and non-destructive technique allowing the simultaneous determination of major chemical constituents. NIR spectroscopy has been used to predict the content of major constituents in sweet potato, cassava, yam and taro crops.[12,30]

This study investigated the potential of NIR spectroscopy as a tool for chemical and textural characterization of *D. alata* tubers. The objectives of the present study were to (1) analyze the chemical and textural characteristics of a panel representative of the genetic diversity of *D. alata*, and (2) assess the potential of NIR spectroscopy as a tool for screening quality attributes.

Materials and methods

Plant material and sample preparation

Twenty-seven *D. alata* accessions (Table 1) and breeding lines from various geographical origins were analyzed. In order to broaden the genetic diversity studied, the accessions were selected from previously obtained genotypic and quality data.[31] Of the selected genotypes, Florido and Bete Bete are the two most cultivated varieties of *D. alata* in Ivory Coast. Bete Bete is considered as a good quality genotype for preparing pounded yam,[32] and was considered as a reference variety throughout this study.

Accessions were planted together in the same plot at the Roujol experimental station (16°10′56″N, 61°35′24″W, 10 a.s.l., Petit-Bourg, Guadeloupe, France) during two cropping seasons (i.e. 2016 and 2017). At harvest, three to five tubers of each variety were peeled with a knife and washed. The head and the tail were removed, and then the tuber was cut longitudinally into two equal parts. The first half was sliced into chips that were oven-dried at 60°C for 72 h and milled into flour using a stainless steel grinder (M-Mandine 150 W, Carrefour, Levallois, France). The granule size was homogenized using a 200 µm sieve. After setting aside 5 g for NIR spectroscopic analysis, samples of 50 g were sent for chemical analysis (i.e. starch, sugar and protein contents at the *Laboratoire d'Analyses Agricoles Teyssier*, Bordeaux, France). In all, 174 yam flour samples (93 flour samples in 2016 and 81 flour samples in 2017) were prepared for further analysis.

The other half tuber was cut into pieces of about 5 mm thick and placed with 1.5 l of water in a pressure cooker. Three identical pressure cookers were used to prepare three samples of each variety at a time. After cooking, the pieces were pounded with a mechanical pounder (Bluesky model, NIF A-2842270, China). The mechanical pounder was used in order to ensure uniform conditions for sample preparation, which could not have been guaranteed with the

Table 1. Details of accessions with their accession code, geographical origin, local name, ploidy level and accession type included in the study.

Code accession	Geographical origin	Local name	Type of accessiona	Ploidy level	Flowering	Genetic diversity groupb
CT133	Vanuatu	Ptris	C	2	Male	3
CT177	Vanuatu	Peter	C	2	Male	6
CT256	Benin	Florido	C	2	Male	7
Plim-G	France (Guadeloupe)	Plimbite	C	2	Male	_
Kin-G	France (Guadeloupe)	Kinabayo	C	2	Female	2
Pac-P	France (Guadeloupe)	Bete Bete	C	2	Female	8
Kab-L	France (Guadeloupe)	Kabusa	C	2	Male	3
PT-IG-00074	Barbados	Oriental	C	2	Female	2
STVB	France (Guadeloupe)	St Vincent blanc	C	2	Female	4
STVV	France (Guadeloupe)	St Vincent violet	C	2	Female	4
Div-PB	France (Guadeloupe)	Divin	C	2	Female	_
PT-IG-00033	USA (Puerto Rico)	Pyramide	C	2	Male	1
CT143	Vanuatu	Malalagi	C	2	Male	3
CT202	Vanuatu	Nureangdan	C	3	Male	14
CT198	Vanuatu	Noulelcae	C	4	Female	16
CT138	Vanuatu	Tagabé	C	4	Female	18
CT148	Vanuatu	Toufi Tetea	C	4	Male	18
160DD	France (Guadeloupe)	Sinoua	C	4	_	_
74F	India	Hyb 2x-74F	BRL	2	Female	_
14M	India	Hyb 2x-14M	BRL	2	Male	_
H4X431	France (Guadeloupe)	Dou	BRL	4	Female	_
H4X105	France (Guadeloupe)	Tiviolet	BRL	4	Female	_
H4X172	France (Guadeloupe)	Hyb 4x-172	BRL	4	Female	_
H4X242	France (Guadeloupe)	Roujol	BRL	4	Male	_
H4X131	France (Guadeloupe)	Hyb 4x-131	BRL	4	Male	_
H4X200	France (Guadeloupe)	Hyb 4x-200	BRL	4	Female	_
H4X274	France (Guadeloupe)	Hyb 4x-274	BRL	4	Male	_

[a]C, Landraces; BRL, Breeding lines.
[b]Genetic groups identified in Arnau et al. (2017).

traditional practice of manual kneading in a mortar with a pestle.[20,33]

Texture evaluation

After pounding, each paste obtained was assessed for its moldability. The moldability of the dough is its ability to form a ball with the hand. Moldability was assigned a binary score depending on whether or not the dough was easily malleable (2: good dough, 1: bad dough).

The textural properties of pounded samples were characterized using texture profile analysis (TPA) with two compression cycles. Samples of yam pastes were formed using a mold (diameter, 25 mm; height, 18 mm) and wrapped in plastic immediately after pounding to prevent surface drying. Nine different pounded yam samples were analyzed for each variety (i.e. three replicates per tuber and three tubers per genotype). TPA was carried out by compressing the yam samples using a TA.Xt2i texture analyser (Stable Micro Systems, Godalming, U.K.). Data were acquired and integrated with Texture Exponent software from the same manufacturer. Each sample was placed horizontally and centered under the probe before measurement. The two compression cycles were achieved using a flat ended aluminum cylinder plunger (i.e. probe P/75; diameter, 50 mm) to 50% of its initial height. The trigger force was 0.049 N and the test speed was set at 1.0 mm s^{-1}. All analyses were carried out at a room temperature of 26°C.

The force deformation curve was evaluated according to Bourne (2002) and Rosenthal (1999).[34,35] Hardness is defined as the maximum force on the first compression cycle. Cohesiveness is defined as a ratio between the positive area under the force-deformation curve of the second and the first compression cycles. Springiness is defined as the ratio between distances from the onset to peak force of the second to the first compression cycle.

Chemical analysis

Dry matter (DM), starch, sugar and protein contents were analyzed according to AFNOR and European Union methods. All measurements were expressed as a percentage of DM (as per standard NF V18-109). Starch was quantified using Ewers protocol (ISO 10,520). Sugars were quantified by the colorimetric method of Luff-Schoorl (CEE 98/54/CE). Total

N content (considered as equivalent total proteins) was calculated using the Kjeldahl method (standard NF V18-100). All analyses were carried out in duplicate with accepted mean coefficients of variation (3% for starch and sugars and 2% for proteins).

Data analysis

The chemical data (starch, proteins and sugars) and TPA texturometer data (hardness, cohesiveness and springiness) were analyzed using XLSTAT version 19.03.44616. A one-way ANOVA was used to analyze the differences in evaluated characteristics between genotypes. A Pearson correlation coefficient analysis was used to determine the correlation coefficients between different chemical and texture characteristics.

The relationship between chemical and TPA parameters was illustrated by a multivariate analysis (principal component analysis, PCA). The results of the PCA were represented by two plots, one for attributes and one for the varieties, with clusters based on hierarchical clustering on principal components (HCPC). The HCPC first built a hierarchical tree. The sums of within-cluster inertia values were calculated for each partition. The partition kept was the one with the highest relative loss of inertia.

Spectra collection and sample selection

NIR spectroscopy analyses were carried out in the food processing laboratory of INRAE's Tropical Animal Research Unit, UR143, in Guadeloupe (France). Two replicates of yam flour samples were scanned with a FOSS-NIRSystems model 6500 scanning monochromator (FOSS-NIRSystems, Silver Spring, MD, USA) equipped with an autocup. The spectroscopic procedures and data recording were conducted with ISIscan(TM) software (FOSS, Hillerød, Denmark). Each flour was placed in a small ring cup 36 mm in diameter, and reflectance spectra from 400 to 2500 nm were recorded at 2 nm intervals. Each spectrum represented the average of 32 scans.[36] Each sample was scanned twice with two independent cups, in order to minimize the effect of particle size. The average spectrum of each sample was used for further chemometric analysis.

Calibration and validation of regression models

For all traits but hardness, modeling was done using WinISI software v4.10.0 (FOSS, NIR spectroscopy, Denmark) using a modified partial least squares regression (M-PLS). First, spectra and data outliers were eliminated following the Shenk and Westerhaus (1993) procedure.[36] Before developing a calibration model, two spectral outlier elimination cycles were set up on the 174 samples using the center algorithm, which calculated the Global H distance (GH) with a cutoff of GH = 3. In this study, an outlier is defined as a sample that does not conform to the bulk of the population in terms of the spectral data. The samples were then divided into a calibration set (3/4) and a validation set (1/4). To select appropriate and representative samples of the calibration set, the SELECT algorithm was applied on the spectra. The number of samples in both calibration and validation set after outlier removal is given in Table 4. A 15-folds cross-validation procedure was implemented to reduce overfitting[36] and split the calibration set into actual calibration and calibration test set allowing identifying the best pretreatments combination and model calibration equation. The validation set was finally used to evaluate the calibration equation. In order to solve problems associated with overlapping peaks and baseline correction, different mathematical pretreatments followed by some pretreatment algorithm were tested. Mathematical treatment parameters were the derivative order (D), the dimension of derivatives (G), the degree of first smooth (S1) and the degree of second smooth (S2). Pretreatment algorithm then tested comprised standard normal variate and detrending (SNV), multi scatter correction (MSC), windowed multi scatter correction (WMSC) and Savitzky-Golay algorithm (SG). The choice of the best pretreatment parameters and combination was based on the standard error of calibration. Mathematical pretreatment ($MP_{D,\ G,\ S1,\ S2}$) and pretreatment algorithm selected were $MP_{2,8,8,1}$ followed by MSC for dry matter, starch and protein; $MP_{2,8,8,2}$ followed by WMSC for sugar; and $MP_{2,4,4,1}$ followed by SNV for springiness and cohesiveness.

Hardness was calibrated with the ChemFlow (https://chemproject.org/ChemFlow) open source software. Data pretreatment used a standard normal variate followed by the Savitzky-Golay algorithm with first order derivative, 13 window size and second degree polynomial. The multivariate distances were used as criteria for removing outliers (i.e. samples in the population that were more variable based on the spectra features). Then, principal component analysis (PCA) was performed followed by the calibration using nonlinear estimation by iterative partial least squares regression (NIPALS),[37] with 4-fold cross validation and 20 latent variables.

At each step (i.e. cross-validation, calibration and validation), the standard error (SE_{CV}, SE_C, SE_P respectively) and the coefficient of determination (R^2_{CV}, R^2_C and R^2_P respectively) were calculated. The best pretreatment and calibration model was selected using the highest R^2_C, and the lowest SE_C.[38] Moreover, the prediction ability of the different regression models was tested based on the ratio between standard deviation (SD) and the standard error (SE_P for validation step). All regression models were developed using the 1100-2498 nm interval range.

Calibration and validation of classification models

In order to evaluate the feasibility of screening genotypes for quality when regression models were performing poorly, continuous texture attributes were binarized based on a variable's average (i.e. values strictly higher than the average were considered to have the traits, while others not). Classification models were then fitted in order to predict belonging to binary classes of hardness, cohesiveness, springiness and moldability.

The classification modeling strategy was based on convolutional neural network (CNN). Recently, machine learning techniques, such as convolutional neural networks (CNNs), have been suggested as a replacement for conventional regression techniques, such as principal component, PLS or support vector machine, due to their superior performance.[39–42] The first advantage of CNN comes when classes are not linearly separable. In such cases the linear classifiers may be outperform by more sophisticated models able to deal with nonlinearity. Moreover, the innovation of convolutional neural networks rely on their ability to automatically learn together a large number of filters specific to a training dataset under the constraints of a specific predictive modeling problem, such as spectra classification.[39] The result is highly specific features that can be detected anywhere on input spectra.

For the calibration of these models, all the spectra were kept without removing spectral outliers. Pretreatments, calibration and validation were carried out using python language (v3.6, https://www.python.org) with a Keras framework (v2.1.5, https://keras.io/) and a TensorFlow backend (v1.6.0, https://www.tensorflow.org). Samples were divided into a calibration set (60) and a validation set (21) using constrained random sampling keeping the same proportion of each class in the two sets.

First, a data (i.e. sample) augmentation was applied to the calibration set only. For each original sample, 30 synthetic spectra were generated using a combination of random translation and rotation of the original spectra. Then spectra presenting absorbance values higher than 1 or lower than 0 were discarded. This allow going from 61 to more than 800 spectra. Noised data augmentation is a common technique used in deep learning to reduce overfitting of small dataset.[43]

Secondly, we made a feature (i.e. spectral) augmentation on calibration and validation sets. We apply an all possibilities approach combining and keeping the different pretreated spectra, including the original one. For each sample, feature augmentation was applied by generating 12 new spectra using pretreatments based on Haar transform, Gaussian derivatives, SVG, SNV, and different degrees of MSC.

A convolutional neural network composed of three convolutional layers followed by two dense layers was fitted to the calibration data. Binary cross entropy was used as the loss function. In order to avoid overfitting, a dropout of 20% of features was applied between layers. The model was calibrated using five-fold cross validation.

Model performance was estimated based on a confusion matrix and traditional sensitivity, specificity, precision, recall, F1 score, Kappa statistic and overall accuracy. Accuracy is a popular metric that refers to the ability of the model to correctly predict the class label of new or unseen data. In addition to this, sensitivity and specificity are also used to assess how well classifiers can recognize true examples as well as false examples. The Kappa statistic evaluates the pairwise agreement between two different observers, corrected for an expected chance agreement. A Kappa value of 0 indicates chance agreement and 1 shows prefect agreement between the classifier and the ground truth (true classes). Precision can be seen as a measurement of exactness or quality, whereas recall is a measurement of completeness or quantity. The F1 score is the harmonic mean of precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.

Results

Chemical evaluation

Significant variation was observed for all the constituents (Table 2). Sugar displayed the largest variation amongst varieties, with a coefficient of variation (CV) of 75%, followed by proteins (CV 17.9%), dry matter (CV 13.7%) and starch (CV 5.5%). The analysis of variance showed that the differences between genotypes were highly significant for all parameters.

Dry matter content ranged from 20.24% to 32.54% (Table 2). The value of control variety Bete Bete (32.54%) was significantly higher than that of control variety Florido (26.78%). In the studied panel, 55% of accessions presented dry matter contents similar to Bete Bete, while eight accessions (30%) presented significantly lower dry matter contents (values between 20 and 24%). Starch content ranged from 66.86% to 82.80%, with five varieties presenting a significantly lower content (i.e. Kinabayo, Oriental, Pyramide, St Vincent violet and St Vincent blanc).

Sugars ranged from 0.52% to 11.6%. The cultivars Oriental, Kinabayo, St Vincent blanc, St Vincent violet and Pyramide had the largest total sugar contents. Proteins ranged from 4.07% to 7.43% with cultivars Pyramide, Kinabayo and Sinoua presenting the highest content.

Textural evaluation

The hardness of the pounded samples ranged from 1.48 N for Sinoua to 11.75 N for Bete Bete.

Table 2. Chemical and textural characteristics of 27 *D. alata* accessions.

Varieties	Dry matter (%)	Starch	Proteins	Sugars	Hardness	Cohesiveness	Springiness	Moldability
14M	29.18 abc	80.61 abcd	5.66 cde	2.05 efhhi	**7.20 bcdef**	0.20 def	0.20 fghi	**2.00 b**
Sinoua	21.07 ef	76.13 f	7.42 a	2.43 efgh	1.48 k	**0.50 ab**	**0.75 ab**	1.00 a
74F	30.14 abc	78.82 bcdef	5.47 cdef	1.24 hij	5.10 ck	0.15 f	0.12 hi	1.00 a
Ptris	30.03 abc	79.19 bcdef	4.95 cdefgh	2.38 efgh	3.57 hijk	0.11 f	0.09 i	1.00 a
Tagabe	31.43 a	80.86 abcd	4.36 fgh	2.37 efgh	**8.13 bcde**	0.20 ef	0.31 defghi	1.33 ab
Malalagi	29.37 abc	81.56 abc	4.11 gh	2.75 defg	5.04 ek	0.16 a	0.18 fghi	1.00 a
Toufi-Tetea	30.16 abc	81.29 abc	4.91 cdefgh	1.41 ghij	**7.49 bcdef**	0.24 def	0.33 defghi	**2.00 b**
Peter	26.91 cd	79.21 bcdef	4.61 efgh	3.87 cd	**7.03 bh**	0.21 def	0.26 efghi	**2.00 b**
Noulelcae	31.55 a	82.80 a	4.23 fgh	1.75 fghij	**7.98 bcdef**	0.22 def	0.26 defghi	**1.67 ab**
Nureangdan	27.13 bcd	78.87 bcdef	5.22 cdefgh	2.46 efgh	5.02 ek	**0.37 bcd**	**0.51 cd**	**2.00 b**
Divin	29.29 abc	77.20 ef	5.67 cde	2.18 efgh	**6.04 bh**	0.15 def	0.15 ghi	1.00 a
Florido	26.78 cd	78.33 cdef	4.40 fgh	3.76 d	5.50 ej	0.21 def	0.32 defghi	1.00 a
Tiviolet	23.10 ef	77.80 def	5.35 cdefg	4.01 cd	4.41 fk	**0.43 abc**	**0.64 bc**	1.00 a
H4X 131	22.74 ef	80.50 abcde	5.85 fgh	0.52 j	2.83 ijk	**0.33 cde**	**0.50 cde**	1.00 a
H4X 172	29.36 abc	80.38 abcde	4.77 defgh	3.07 def	**7.84 bcdef**	0.23 def	0.35 defgh	1.33 ab
H4X 200	31.31 a	82.66 a	4.83 defgh	1.20 hij	**9.02 abc**	0.25 def	0.37 defg	**1.67 ab**
Roujol	29.80 abc	81.31 abc	4.51 efgh	2.04 efghi	**9.77 ab**	0.23 def	0.39 defg	**2.00 b**
H4X 274	31.00 ab	80.76 abcd	5.01 cdefgh	2.49 efgh	**8.10 bcde**	0.27 def	0.39 defg	1.33 ab
Dou	31.46 a	82.00 ab	4.30 fgh	2.22 abcd	**8.68 abcd**	**0.33 cde**	0.39 def	**2.00 b**
Kabusa	31.29 a	78.61 bcdef	4.95 cdefgh	3.33 de	1.73 k	**0.46 abc**	**0.71 abc**	1.00 a
Kinabayo	21.38 ef	66.86 h	7.01 ab	7.36 b	2.20 jk	**0.45 abc**	**0.63 bc**	1.00 a
Oriental	23.69 def	70.41 g	4.07 h	11.69 a	2.72 ijk	**0.53 a**	**0.63 bc**	1.00 a
Bete Bete	32.54 a	80.54 abcde	5.28 cdefgh	2.17 efgh	**11.75 a**	0.22 def	0.32 defghi	**2.00 b**
Plimbite	28.70 abc	78.71 bcdef	6.12 bc	0.81 ij	3.85 ghijk	0.24 def	0.30 defghi	1.00 a
Pyramide	20.24 f	70.06 g	7.43 a	5.01 c	4.41 fk	**0.46 abc**	**0.65 bc**	1.00 a
St Vincent blanc	22.80 ef	70.84 g	5.42 cdef	6.87 b	2.75 ijk	**0.57 a**	**0.90 a**	1.00 a
St Vincent violet	24.48 de	71.46 g	6.14 bc	6.28 b	3.70 gk	**0.46 abc**	**0.74 ab**	1.00 a
Maximum	32.54	82.80	7.43	11.69	11.75	0.57	0.90	2.00
Minimum	20.24	66.86	4.07	0.52	1.48	0.11	0.09	1.00
Mean	27.66	78.07	5.26	3.25	5.68	0.30	0.42	1.35
Standard deviation	3.79	4.32	0.94	2.42	2.70	0.13	0.22	0.44
Coefficient of variation (%)	13.70	5.53	17.79	74.44	47.60	44.12	51.41	32.59
p-Value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Different letters indicate statistically significant differences between genotypes.
Values in bold represent a binary classification of texture attributes. If a value is strictly higher than the attribute average, the genotype is considered as having the respective texture attribute (e.g. hard for hardness or cohesive for cohesiveness).

Amongst the accessions, 44% presented hardness values similar to Bete Bete (not significantly different), while others presented significantly ($p < 0.05$) lower values, including the control variety Florido. Cohesiveness ranged from 0.11 to 0.57. Bete Bete and Florido presented very similar cohesiveness values (0.22 and 0.21). Eight accessions presented significantly ($p < 0.05$) higher cohesiveness values than the control varieties. Springiness ranged from 0.09 to 0.90. Bete Bete and Florido presented identical springiness values (0.32). Eight accessions presented significantly higher ($p < 0.05$) springiness values than the control varieties.

Moldability ranged from 2 to 1 with seven genotypes regarded as moldable (i.e. 14 M, Toufi Tetea, Peter, Noulelcae, H4X200, Roujol, Dou and Bete Bete). One of the two control varieties (i.e. Florido) was characterized as not moldable.

Relationships between main variables

The correlation analysis (Table 3) revealed very highly significant linear correlations between the main chemical parameters analyzed. Dry matter was positively correlated with starch and negatively correlated with protein and sugar contents. There was also a very highly significant linear correlation between the five texture parameters analyzed. Springiness was positively correlated with cohesiveness ($R^2 = 0.97$). Hardness was negatively correlated with springiness and cohesiveness.

The correlation coefficients between chemical and textural parameters revealed some significant relationships. Dry matter and starch were both positively correlated with hardness and negatively correlated with springiness and cohesiveness. Sugars and proteins were both positively correlated with cohesiveness and springiness.

Table 3. Table of correlations (r) for the main continuous variables.

Variables	1	2	3	4	5	6
1. Dry matter	-					
2. Starch	0.762***[a]	-				
3. Proteins	−0.592***	−0.559***	-			
4. Sugars	−0.541***	−0.807***	0.061 ns	-		
5. Hardness	0.638***	0.584***	−0.416**	−0.401**	-	
6. Cohesiveness	−0.673***	−0.641***	0.344**	0.601***	−0.503***	-
7. Springiness	−0.656***	−0.598***	0.354**	0.527***	−0.476***	0.958***

[a]Level of significance: ns (not significant).
*$p < 0.05$; **$p < 0.01$; ***$p < 0.0001$.

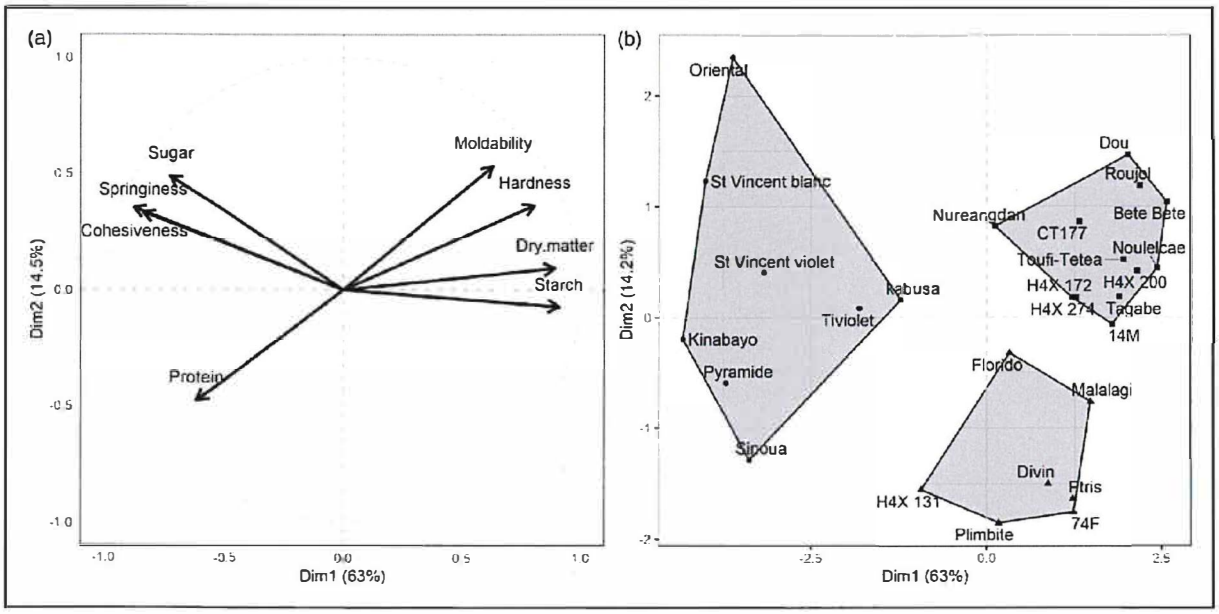


Figure 1. Principal component analysis of the main variables (a) and hierarchical clustering on principal components for the 27 individuals (b).

The principal component analysis conducted on the data matrix indicated the respective contribution of the seven variables to the projection, with dimensions 1 and 2 explaining respectively 62.97 and 14.54% of total variance (Figure 1). Dimension 1 was negatively linked with the attributes cohesiveness ($r = -0.87$), springiness ($r = -0.83$), sugars ($r = -0.72$) and proteins ($r = -0.62$), while it was positively linked with starch ($r = 0.91$), dry matter ($r = 0.90$), hardness ($r = 0.80$) and moldability ($r = 0.63$). The positive side of the second dimension was primarily linked with moldability ($r = 0.56$). The protein content was also linked with the negative side of the third dimension ($r = 0.56$).

The hierarchical clustering on principal components revealed three main groups of varieties at the extremities of each dimension. The group including the control genotype Bete Bete was characterized by high dry matter, starch content and hardness, and good moldability. The group including the second control (i.e. Florido) differed from the first by a higher protein content and lower hardness and moldability. Finally, the last group contained genotypes with high sugar and protein contents, high cohesiveness and springiness, and low moldability.

Figure 2 shows the results of the principal component coordinates for moldable and non-moldable individuals. The upper left window of the plot groups a mix of non-moldable samples and all the moldable samples but one. On the other hand, the other windows were mainly populated with non-moldable individuals.

Calibration and validation of predictive models

The results in Table 4 show the calibration performance for chemical and textural attributes. The chemical attribute models performed well during the calibration step ($R^2_C > 0.84$ and $R^2_{CV} > 0.79$). On the other hand, all the texture parameters showed R^2_C values below 0.8, except hardness (0.83). The good performance of the chemical models was

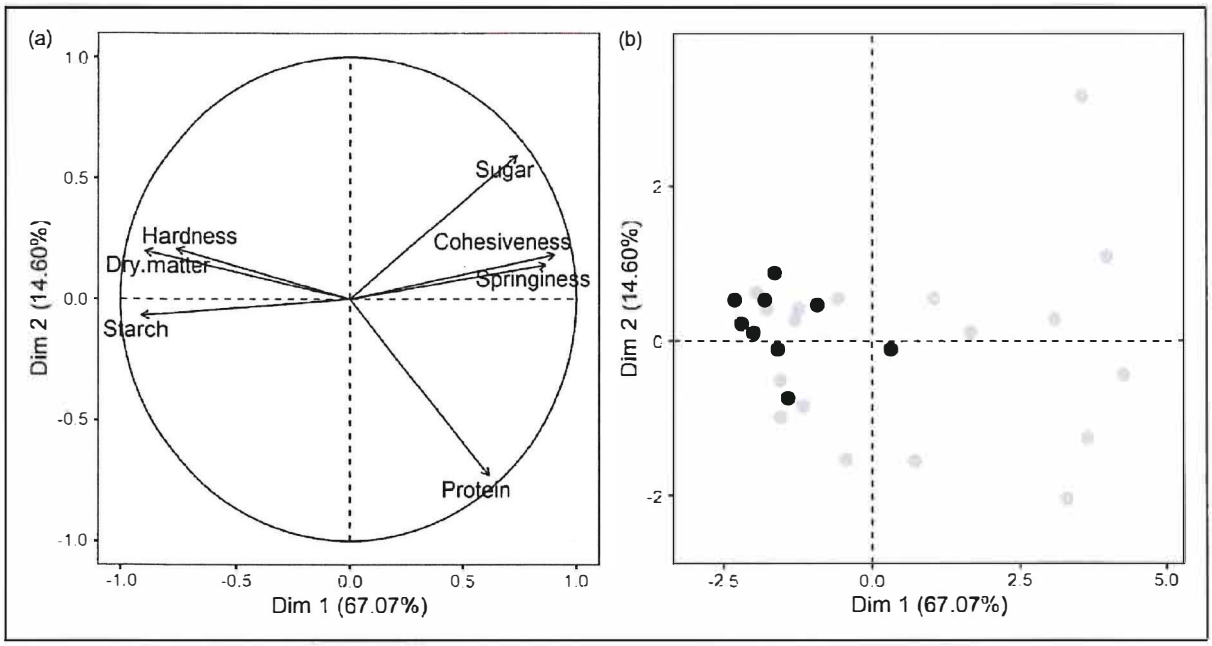


Figure 2. Principal component analysis of the main variables (a) and principal component coordinates for moldable (black) and nonmoldable (gray) individuals (b).

Table 4. Model performance metrics for the calibration (C), cross-validation (CV), and prediction (P) steps.

	Procedure			Calibration			Cross-validation		Validation			
Constituent	Model	Latent variables	SEL	N_C	SE_C	R^2_C	SE_{CV}	R^2_{CV}	N_P	SE_P	R^2_P	SD/SE_P
Dry matter (%)	MPLS	15	1	141	1.44	0.84	1.64	0.79	37	1.58	0.85	2.13
Protein (%)	MPLS	15	2	136	0.18	0.96	0.24	0.92	37	0.29	0.88	2.77
Sugar (%)	MPLS	9	3	136	0.32	0.96	0.42	0.93	38	0.56	0.93	3.49
Starch (%)	MPLS	15	3	146	0.91	0.91	1.21	0.83	37	1.46	0.89	2.67
Hardness (*N*)	NIPALS	20	1	54	1.66	0.83	2.17	0.73	19	1.68	0.58	1.68
Cohesiveness (dimensionless)	MPLS	9	1	52	0.08	0.55	0.10	0.23	21	0.11	0.55	1.35
Springiness (dimensionless)	MPLS	9	1	54	0.13	0.57	0.16	0.41	21	0.19	0.52	1.24

SEL: standard error of laboratory reference measurement; N: number of samples; SE: standard error; R^2: coefficient of determination; SD: standard deviation; MPLS: modified partial least square regression model; NIPALS: nonlinear estimation by Iterative partial least squares regression.

confirmed when the model was validated against the validation data set (Figure 3, $R^2_P > 0.85$). The performance to deviation ratio (SD/SE_P) showed that the model for dry matter, protein, sugar and starch content, with values higher than two, could be considered good for screening purposes.[44,45] However, the calibration performance for hardness ($R^2_P = 0.58$, $SD/SE_P = 1{,}68$), cohesiveness ($R^2_P = 0.55$, $SD/SE_P = 1.35$) and springiness ($R^2_P = 0.52$, $SD/SE_P = 1.24$) displayed a poor predictive performance (Figure 3).

Figure 4 presents the confusion matrix and the model performance metrics for the different classification models during the cross-validation (A) and validation (B) steps. The convolutional neural networks exhibited high accuracy (>0.8), except for hardness (0.55), where the model lacked specificity (0.25). For moldability, the model was characterized by very good sensitivity (1), but rather poor specificity (0.636). Springiness and cohesiveness where classified with very good performance metrics.

Discussion

All the chemical and texture attributes studied were highly variable and their differences between genotypes were highly significant. Among these traits, dry matter and starch, two strongly and positively correlated parameters, are known to affect yam quality.[15,21] This study showed that dry matter and starch were both negatively correlated with sugar and protein contents. These results tallied with those obtained on root and tuber plants: sweet potato, cassava, taro, potato and yam (i.e. *D. alata*, *D. bulbifera*, *D. cayenensis-rotundata*, *D. esculenta*, *D. nummularia*,

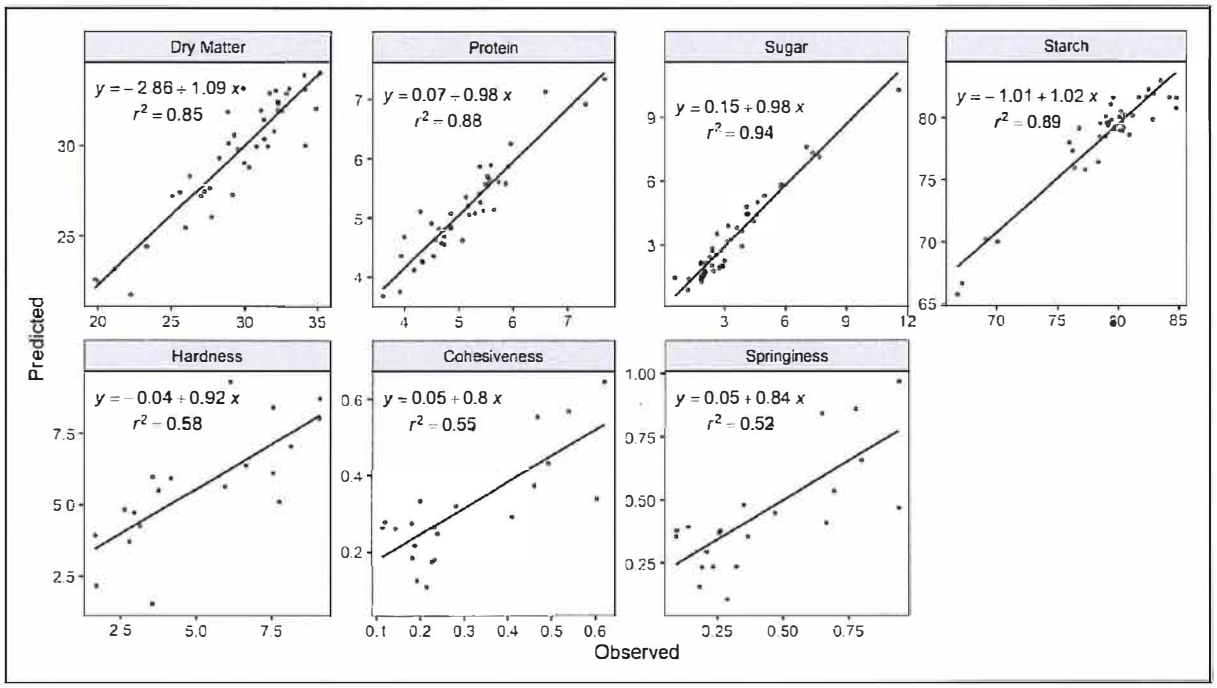


Figure 3. Comparison of observed and predicted values for chemical ($n = 41$) and textural ($n = 22$) attributes during the validation step.

D. pentaphylla and *D. transversa*).[15,46] However, on a panel of 48 varieties within the species *D. alata*, Lebot et al, (2005) showed that dry matter was positively correlated with sugar.[15] This discrepancy may be due to the longer storage period applied in this study, leading to the breaking of dormancy and starch remobilization to produce sugars. In this particular situation, the sugar content could be negatively correlated to the starch content. Bete Bete, a control variety known for its gustatory qualities, as well as 55% of the varieties evaluated, had significantly higher dry matter content than the others. Although all but one of the moldable genotypes belonged to the high dry matter varieties, only half of these high dry matter varieties were evaluated as moldable. This result indicated that if dry matter content impacts sensory quality it may be a necessary, but not sufficient, attribute interacting with other traits. Conversely, varieties with the highest sugar (i.e. Oriental, Kinabayo, St Vincent blanc, St Vincent violet and Dou) and protein (i.e. Pyramide, Sinoua, Kinabayo, St Vincent violet and Plimbite) contents were all evaluated as non-moldable.

The texture parameters also displayed a high degree of variability. Hardness is one of the most important texture parameters in product evaluation.[34,35] Brunnschweiler et al., (2005) showed that varieties of high quality *D. rotundata* were much firmer than those of *D. alata*.[47] In this study, all the moldable genotypes belonged to varieties classified as hard. But this class also contained some non-moldable genotypes. This suggests that there could be a hardness threshold below which yam can no longer be moldable. Conversely, high hardness does not automatically imply that the product will be of high quality. Hardness threshold identification could be an important criterion for selecting *D. alata* varieties.

Cohesiveness is the degree of malleability of a product. It is the way in which a product or paste is easily molded with the fingers into a ball or spherical shape.[47] In this study, eleven varieties presented high cohesiveness values. Amongst them, only one (Dou) was moldable. The varieties with low cohesiveness values belonged to both the moldable and non-moldable categories. Again, the result suggested a cohesiveness threshold above which yams are no longer moldable. On the other hand, low cohesiveness does not automatically imply that the product will be moldable.

Otegbayo *et al.* (2007) showed that difficult-to-mold *D. alata* varieties have high elasticity values.[27] This is in accordance with the high correlation found between springiness and cohesiveness. All the springy genotypes where not moldable, while the rigid ones included both types (i.e. moldable and non-moldable).

The models for the quantitative prediction of physico-chemical parameters (i.e. dry matter, proteins, starch and sugars) showed a good performance, unlike the texture parameters (i.e. hardness, cohesiveness, springiness). These results are consistent with previous studies which found a good calibration performance for dry matter, starch, protein and sugars on different yam, taro and cassava species.[12,46,48,49] The relatively small number of samples available for texture traits calibration may result in reduced

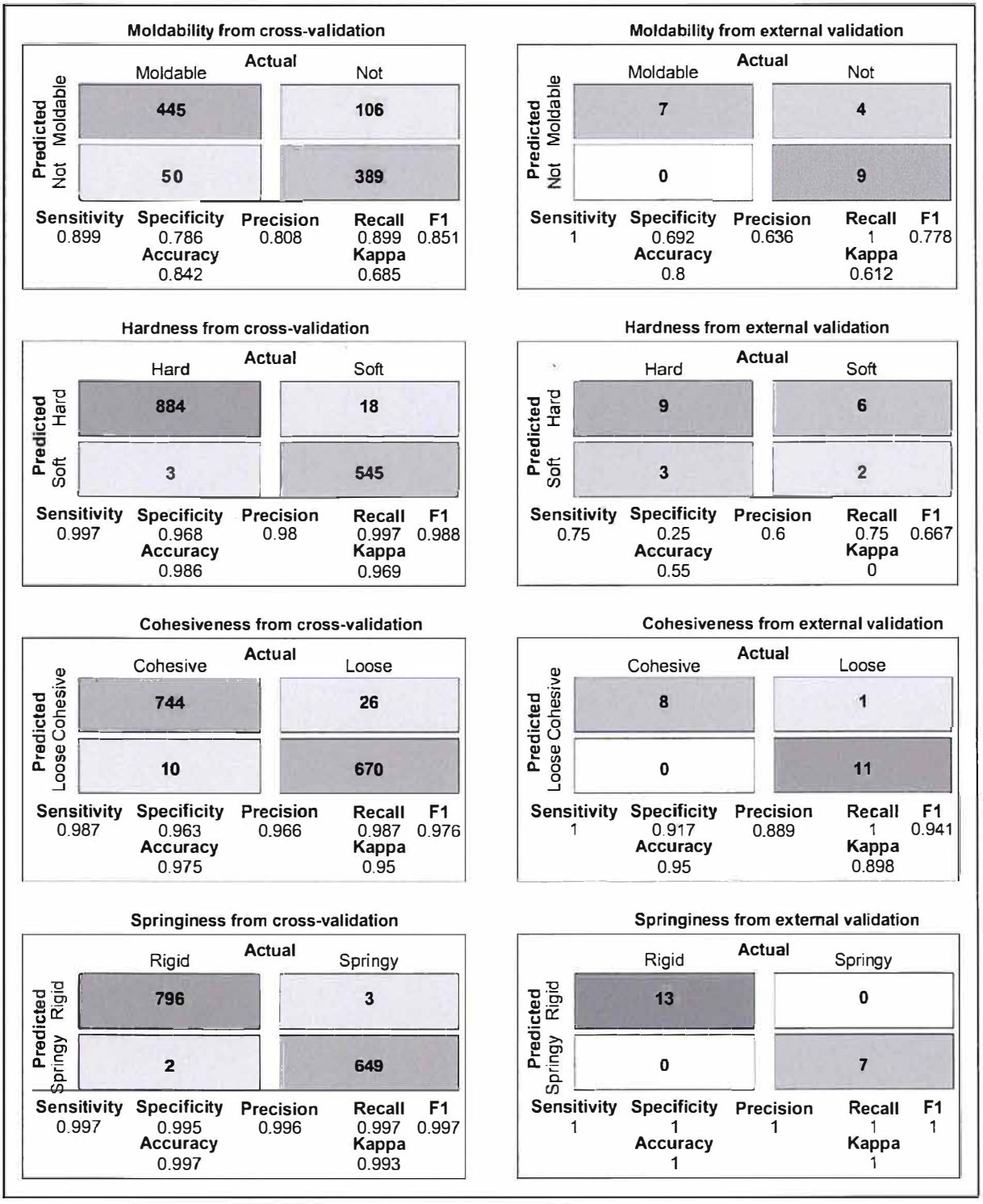


Figure 4. Confusion matrix, sensitivity, specificity, precision, recall, F1-measure, accuracy and Kappa data for yam texture classification models during cross-validation (left plots) and validation (right plots) steps.

performance and robustness. Current prediction models should benefit from supplementary calibration data and external validation.

However, while regression models failed to predict texture parameters, classification models provided an accurate tool for qualitative screening varieties for their moldability, springiness and cohesiveness. Only the hardness classification model still exhibit poor performance and seem to be prone to overfitting. Indeed, accuracy decrease between calibration and validation steps (i.e. 0.99 to 0.55) suggest such an overfitting.

Conclusion

Despite high variability in the main chemical and texture variables studied, it was not possible to define a common profile discriminating between moldable and non-moldable individuals. However, although the results failed to identify a precise profile for moldable genotypes, they enabled us to ascertain what leads to a lack of moldability: high protein content, low hardness, or high cohesiveness. These minimum requirements could be very useful prerequisites for breeders to screen and discard low quality genotypes.

Moreover, the results confirmed that near infrared spectrometry could be used for rapid screening of dry matter, protein, sugar and starch contents in *D. alata* yam varieties. Although texture parameters could not be satisfactorily quantitatively predicted, the classification algorithm proved to be accurate for qualitative prediction. For breeding purposes, using thresholds or classification algorithms to screen for texture quality traits seems to be a good compromise between precision and speed. Using near infrared spectra with binary classification modeling allows for high-throughput phenotyping of moldability, cohesiveness and springiness.

Acknowledgements

We acknowledge the CRB-PT biological resource center (INRA/CIRAD, France) for providing plant material samples. We would also like to thank Jocelyne Leinster, Pascale Bade and Valerius Calif for their technical assistance for NIR spectroscopy acquisition. English Language editing and review services supplied by Peter Biggins.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was made possible by the RTBfood project funded by the Bill and Melinda Gates Foundation [OPP1178942], for which we should like to thank the various leaders. The authors acknowledge funding provided by the ERDF Cavalbio project, financed by *Région Guadeloupe* and the European Union, which contributed to field preparation and maintenance.

ORCID iDs

Denis Cornet https://orcid.org/0000-0001-9297-2680
Amani Michel Kouakou https://orcid.org/0000-0001-7232-0554

References

1. Coursey DG. Yams. London: Longmans, 1967.
2. Cornet D, Sierra J and Bonhomme R. Characterization of the photosynthetic pathway of some tropical food yams (*Dioscorea* spp.) using leaf natural 13C abundance. *Photosynthetica* 2007; 45: 303–305.
3. Abraham K and Nair PG. Floral biology and artificial pollination in *Dioscorea alata* L. *Euphytica* 1990; 48: 45–51.
4. Mignouna D, Mank A, Ellis N, et al. A genetic linkage map of water yam (*Dioscorea alata* L.) based on AFLP markers and QTL analysis for anthracnose resistance. *Theor Appl Genet* 2002; 105: 726–735.
5. Cornet D, Sierra J, Tournebize R, et al. Bayesian network modeling of early growth stages explains yam interplant yield variability and allows for agronomic improvements in West Africa. *Eur J Agron* 2016; 75: 80–88.
6. FAOTSTAT 2018. www.fao.org/faostat/en/#data/FO (accessed 18 March 2020).
7. Orkwor GC, Asiedu R and Ekanayake IJ. *Food yams: advances in research*. Umudike, Nigeria: IITA and NRCRI, 1997.
8. Egesi NC, Asiedu R, Egunjobi JK, et al. Genetic diversity of organoleptic properties in water yam (*Dioscorea alata* L.). *J Sci Food Agric* 2003; 83: 858–865.
9. Doumbia S, Toure M and Mahyao A. Commercialisation de l'igname en côte d'Ivoire, état actuel et perspectives d'évolution. *Cahiers Agricultures* 2006; 15: 273–277.
10. Egesi CN and Asiedu R. Analysis of yam yields using the additive main effects and mutiplicative interaction (AMMI) model. *Afr Crop Sci J* 2002; 10: 195–201.
11. Ano G, Anaïs G and Chidiac A. Creation et utilisation de varietes resistantes aux maladies, elements essentiels de la diversification agricole en Guadeloupe. *Phytoma* 2002; 551: 36–37.
12. Lebot V, Champagne A, Malapa R, et al. NIR determination of major constituents in tropical root and tuber crop flours. *J Agric Food Chem* 2009; 57: 10539–10547.
13. Arnau G, Nemorin A, Maledon E, et al. Advances on polyploid breeding in yam *D. alata*. *In: Proceedings of the first international symposium on roots, rhizomes, tubers, plantains, bananas and papaya*, Santa Clara, Cuba, 7–10 November 2011.
14. Martin FW. Tropical yams and their potential. Part. 2: *Dioscorea bulbifera*. *USDA Agric Handb* 1974; 466: 1–19.
15. Lebot V, Malapa R, Molisale T, et al. Physico-chemical characterisation of yam (*Dioscorea alata* L) tubers from Vauatu. *Genet Resour Crop Evol* 2006; 53: 1199–1208.
16. Akissoe HN, Mestres C, Hounhouigan DJ, et al. Sensory and psychochemical quality quality of pounded yam: Varietal and storage effects. *J Food Process Preserv* 2009; 33: 75–90.
17. Dufie BF. *Characterizaion of water yam (Dioscorea alata) for existing and potential food products*. PhD Thesis, Kwame Nkrumah University of Science and Technology, Ghana, 2009.
18. Brunnschweiler J, Mang D, Farah Z, et al. Structure–texture relationships of fresh pastes prepared from different yam (*Dioscorea* spp) varieties. *LWT – Food Sci Technol* 2006; 39: 762–769.
19. Otegbayo B, Aina J, Asiedu R, et al. Microstructure of boiled yam (*Dioscorea* spp) and its implication for assessment of textural quality. *J Texture Stud* 2005; 36: 324–332.

20. Nindjin C, Otokoré D, Hauser S, et al. Determination of relevant sensory properties of pounded yams (*Dioscorea* spp.) using a locally based descriptive analysis methodology. *Food Qual Prefer* 2007; 18: 450–459.
21. Otegbayo B, Asiedu R and Bokanga M. Physicochemical properties of yam starch: effect on textural quality of yam food product (pounded yam). *J Food Agric Environ* 2011; 9: 145–150.
22. Friedman HH, Whitney JE and Szczesniak AS. The texturometer – a new instrument for objective texture measurement. *J Food Sci* 1963; 28: 390–396.
23. Yasumatsu K, Sawada K, Moritaka S, et al. Whipping and emulsifying properties of soybean products. *Agr Biol Chem* 1972; 36: 719–727.
24. Gibert O, Giraldo A, Uclés-Santos JR, et al. A kinetic approach to textural changes of different banana genotypes (*Musa* sp.) cooked in boiling water in relation to starch gelatinization. *J. Food Eng* 2010; 98: 471–479.
25. Sajeev MS, Sreekumar J, Unnikrishnan M, et al. Kinetics of thermal softening of cassava tubers and rheological modeling of the starch. *J Food Sci Technol* 2010; 47: 507–518.
26. Polycarp D. Textural characteristics of seven different yams (*Dioscorea* species) grown and consumed in Ghana. *Int J Sci Technol Res* 2010; 6: 281–284.
27. Otegbayo B, Aina J, Abbey L, et al. Texture profile analysis applied to pounded yam. *J Texture Stud* 2007; 38: 355–372.
28. Huang Y, Xu F, Hu H, et al. Development of a predictive model to determine potato flour content in potato-wheat blended powders using near-infrared spectroscopy. *Int J Food Prop* 2018; 21: 2030–2036.
29. Huang HB, Yu HY, Xu HR, et al. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: a review. *J. Food Eng* 2008; 87: 303–313.
30. Zum Felde T, Burgos G, Espinoza J, et al. Screening for b-carotène, iron, zinc, starch, individual sugars and protein in sweetpotato germplasm by Near-Infrared Reflectance. In: *15th triennial symposium of the international society for tropical root crops*, Lima, Peru, 2009.
31. Arnau G, Bhattacharjee R, Mn S, et al. Understanding the genetic diversity and population structure of yam (*Dioscorea al*ata L.) using microsatellite markers. *PLoS One* 2017; 12: e0174150. doi.org/10.1371/journal.pone.0174150.
32. Brunnschweiler J. *Structure and texture of yam (Dioscorea spp.) and processed yam products*. PhD Thesis, Swiss Federal Institute of Technology, Switzerland, 2004.
33. Bricas N, Vernier P, Ategbo E, et al. The development of pounded yam machine in West Africa. *J Res Dev* 1997; 44: 100–114.
34. Bourne MC. *Food texture and viscosity: concept and measurement*. San Diego, CA: Academic Press, 2002.
35. Rosenthal AJ. *Food texture: Measurement and perception*. Aspen: Springer; 1999.
36. Shenk JS and Westerhaus MO. *Analysis of agriculture and food products by near-infrared reflectance spectroscopy*. Port Matilda, PA: Infrasoft, 1993.
37. Rossard V, Boulet JC, Gogé F, et al. ChemFlow, chemometrics using galaxy. *F1000Research* 2016; 5: 1671.
38. Wu JG, Shi CH and Zhang XM. Estimating the amino acid composition in the milled rice powder by near-infrared reflectance spectroscopy. *Field Crops Res* 2002; 75: 1–7.
39. Cui C and Fearn T. Modern practical convolutional neural networks for multivariate regression: applications to NIR calibration. *Chemometr Intell Lab* 2018; 182: 9–20.
40. Sarin JK, t, Moller NCR, Mohammadi A, et al. Machine learning augmented near-infrared spectroscopy: *in vivo* follow-up of cartilage defects. *Osteoarthr Cartil* 2020. DOI: 10.1016/j.joca.2020.12.007.
41. Yang SY, Kwon O, Park Y, et al. Application of neural networks for classifying softwood species using near infrared spectroscopy. *J Near Infrared Spectrosc* 2020; 28: 5–6.
42. Zhang L, Xiangqian D and Hou R. Classification modeling method for near-Infrared spectroscopy of tobacco based on multimodal convolution neural networks. *J Anal Methods Chem* 2020; 2020: 9652470.
43. Bae HJ, Kim CW, Kim N, et al. A Perlin noise-based augmentation strategy for deep learning with small data samples of HRCT images. *Sci Rep* 2018; 8.
44. Lebot V and Malapa R. Application of near infrared reflectance spectroscopy to the evaluation of yam (*Dioscorea alata*) germplasm and breeding lines. *J Sci Food Agric* 2013; 93: 1788–1797.
45. Roberts CA, Workman J and Reeves JB. *Near-infrared spectroscopy in agriculture*. Agronomy Monographs Volume 44. Madison, WI: ASA, CSSA, SSSA, 2004.
46. Williams PC. Implementation of near-infrared technology. In: Williams P and Norris K (eds), *Near-infrared technology in the agriculture and food industries*. 2nd ed. St. Paul, MN: American Association of Cereal Chemists, 2001, pp. 145–169.
47. Brunnschweiler J, Luethi D, Handschin S, et al. Isolation, physicochemical characterization and application of yam (*Dioscorea* spp.) starch as thickening and gelling agent. *Starch/Stärke* 2005; 57: 107–117.
48. Lebot V, Malapa R and Jung M. Use of NIRS for the rapid prediction of total N, minerals, sugars and starch in tropical root and tuber crops. *NZ J Crop Hort Sci* 2013; 41: 144–153.
49. Ikeogu UN, Davrieux F, Dufour D, et al. Rapid analyses of dry matter content and carotenoids in fresh cassava roots using a portable visible and near infrared spectrometer (Vis/NIRS). *PLoS One* 2017; 12: e0188918.

Publication 5

• Arnau, G., Desfontaines, L., Ehounou, A. E., Marie-Magdeleine, C., Kouakou, A. M., Leinster, J., Nudol, E., Maledon, E., & Chair, H. (2024). Quantitative trait loci and candidate genes for physico-chemical traits related to tuber quality in greater yam (*Dioscorea alata* L.). *Journal of the Science of Food and Agriculture*, 104(8), 4872-4879.

Research Article

Received: 18 January 2023 Revised: 16 May 2023 Accepted article published: 4 July 2023 Published online in Wiley Online Library: 29 July 2023

(wileyonlinelibrary.com) DOI 10.1002/jsfa.12822

Quantitative trait loci and candidate genes for physico-chemical traits related to tuber quality in greater yam (*Dioscorea alata* L.)

Gemma Arnau,[a,b*] Lucienne Desfontaines,[c] Adou Emmanuel Ehounou,[d] Carine Marie-Magdeleine,[e] Amani Michel Kouakou,[d] Jocelyne Leinster,[c] Elie Nudol,[b,f] Erick Maledon[b,f] and Hana Chair[a,b]

Abstract

BACKGROUND: Starch, dry matter content (DMC), proteins, and sugars are among the major influences on yam tuber quality. Genetic improvement programs need simple, rapid, and low-cost tools to screen large populations. The objectives of this work were, using a quantitative trait loci mapping approach (QTL) on two diploid full-sib segregating populations, (i) to acquire knowledge about the genetic control of these traits; (ii) to identify markers linked to the genomic regions controlling each trait, which are useful for marker-assisted selection (MAS); (iii) to validate the QTLs on a diversity panel; and (iv) to identify candidate genes from the validated QTLs.

RESULTS: Heritability for all traits was moderately high to high. Significant correlations were observed between traits. A total of 25 QTLs were identified, including six for DMC, six for sugars, six for proteins, and seven for starch. The phenotypic variance explained by individual QTLs ranged from 14.3% to 28.6%. The majority of QTLs were validated on a diversity panel, showing that they are not specific to the genetic background of the progenitors. The approximate physical location of validated QTLs allowed the identification of candidate genes for all studied traits. Those detected for starch content were mainly enzymes involved in starch and sucrose metabolism, whereas those detected for sugars were mainly involved in respiration and glycolysis.

CONCLUSION: The validated QTLs will be useful for breeding programs using MAS to improve the quality of yam tubers. The putative genes should be useful in providing a better understanding of the physiological and molecular basis of these important tuber quality traits.

Keywords: dry matter content; starch biosynthesis; molecular breeding; heritability

INTRODUCTION

Yams (*Dioscorea* spp.) are herbaceous vines cultivated for their starchy tubers. They represent a major food crop in the tropical and subtropical regions of Oceania, Asia, the Caribbean, South America, and in particular in West Africa, which accounts for 92% of the world's production.[1] *Dioscorea rotundata* and *D. alata* are the most important cultivated species. *Dioscorea alata* ranks second in production importance, but it is the world's most widely distributed species. This is due to the fact that it offers several particular advantages in terms of early vigor for weed control, yield potential under low to medium soil fertility conditions, and a better tuber storage ability.[2,3]

In West Africa, yams are consumed in several ways (boiled, fried, roasted, baked, or pounded, after being cooked, into a stiff paste called *fufu*). Cultivars of *D. rotundata* are more suitable for the preparation of *fufu*. However one study showed that tubers from

* *Correspondence to: G Arnau, CIRAD Montpellier Centre, UMR Agap Institut, F-34398 Montpellier, France. E-mail: gemma.arnau@cirad.fr*

a *CIRAD, UMR AGAP Institut, Montpellier, France*

b *UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France*

c *INRAE, UR 1321 ASTRO Agrosystèmes tropicaux. Centre de recherche Antilles-Guyane, Petit-Bourg, France*

d *CNRA, Station de Recherche sur les Cultures Vivrières (SRCV), Bouaké, Côte d'Ivoire*

e *INRAE, ASSET, Petit-Bourg, France*

f *CIRAD, UMR AGAP Institut, Petit Bourg, France*

some *D. alata* genotypes could form a good dough, comparable to that of some *D. rotundata* genotypes.[4]

The sensory quality of tubers depends on many physico-chemical and textural characteristics.[5,6] It has been shown that the quality of tubers for consumption (either boiled or pounded) was strongly related to their dry matter content (DMC), starch, and amylopectin content.[7] The four main constituents of tubers are starch, DMC, sugars, and proteins. Tubers of *D. alata* contain 20–40% of DMC, 60–80% starch, 0.5–11.6% sugars, and 4.0–7.4% proteins.[8]

Genetic improvement programs need simple, rapid, and low-cost tools to screen large populations. Near-infrared reflectance spectroscopy (NIRS) has been demonstrated to be a reliable technique to predict the major tuber constituents in *D. alata* yam species.[9,10] However, as spectra were generated from flour samples and not from raw samples, this protocol requires a long sample-processing time and remains rather difficult to apply to a large number of genotypes.

Marker-assisted selection could be a high-throughput method facilitating breeding efforts. Indeed, with the development of new-generation sequencing technologies, it has become much easier to search for genomic regions associated with traits of interest. A few studies have been conducted with yam to elucidate the genetic determinism of tuber quality related traits. By using a quantitative trait loci (QTL) mapping approach on two biparental populations, several genomic regions linked to important morphological and agronomic tuber quality traits have been identified.[11] The heritability of DMC was estimated on a diversity panel including eight different *Dioscorea* species.[12] A Genome-Wide-Association study was carried out in *D. alata* and some single-nucleotide polymorphism (SNP) markers linked to DMC could be identified.[13]

The current study used a QTL mapping approach on two diploid full-sib segregating populations, aiming: (i) to acquire knowledge about the genetic control of starch, DMC, sugar, and protein content in tubers; (ii) to identify markers linked to the genomic regions controlling each trait, useful for marker-assisted selection (MAS); (iii) to validate the QTLs on a diversity panel; and (iv) to identify candidate genes from the validated QTLs.

MATERIALS AND METHODS

Plant material and sample preparation

Two diploid full-sib segregating *D. alata* populations composed of 93 (population A, 74F x Kabusa) and 80 (population B, 74F x 14 M) progenies, respectively, were used to map the quality traits. Both populations were generated previously[11] and derived from crosses between diploid progenitors contrasting for expression of quality traits. The male Kabusa and 14 M parents have significantly higher DMC and starch content than the female parent 74F, which has, on the other hand, higher sugar content than 14 M and lower sugar content than Kabusa. Progenitors do not differ significantly in protein content.

The mapping populations and progenitors were planted in two blocks at Roujol experimental station in Guadeloupe, France (16° 10′ 56″ N, 61° 35′ 24″ W, 10 m.a.s.l.) during cropping season 2017–2018. Each block included nine plants of each genotype. After harvest, about 200 g of the central part of three to four tubers per genotype was sliced into chips and dried in an oven at 60 °C for 72 h (3 to 4 tubers × 200 g × 2 blocks). Then flours of each tuber were prepared for NIRS screening as described in Ehounou *et al*. (2021).[10]

The diversity panel used for the validation of QTLs consisted of 24 *D. alata* genotypes and included landraces and breeding lines presenting a high diversity for the four traits whose qualities were studied. These accessions were planted together in the same field during two cropping seasons (2016–2017 and 2017–2018) in Guadeloupe (16° 10′ 56″ N, 61° 35′ 24″ W, 10 m.a.s.l.). After harvest, as for biparental populations, about 200 g of the central part of three to five tubers of each genotype were sliced into chips and oven dried at 60 °C for 72 h (3 to 5 tubers × 200 g × 2 years). Then flour from each tuber was prepared for chemical analysis as described in Ehounou *et al*. (2021).[10]

Phenotyping of progenies and the diversity panel

Phenotyping of 24 accessions of the diversity panel was carried out using the chemical analysis and methods described in Ehounou *et al*. (2021).[10] Starch, sugar, and protein content was expressed as a percentage of dry weight.

Dry matter content, starch, sugar, and protein content of progenies and progenitors were predicted by NIRS analysis, using a FOSS-NIRSystems model 6500 scanning monochromator (FOSS-NIRSystems, Silver Spring, MD, USA) and partial least squares models developed previously.[10] Two replicates were scanned for each flour sample. A total of 1762 NIRS measures were carried out in the technical platform of INRAE's UR143 ASSET research unit, in Guadeloupe. The NIRS spectra were generated from a single block for population B.

Statistical analysis

Pearson correlation tests, histograms, box plots, ANOVA, and normality analysis were performed using XLSTAT version 19.03.44616. Distributions of progeny phenotypic data were tested for normality using Shapiro–Wilk and JarqueBera tests. Broad-sense heritability was estimated as described in Ehounou *et al*. (2022).[11]

Genotyping by sequencing

Single-nucleotide polymorphism genotyping data of both mapping populations, progenitors and the diversity panel were generated previously.[11,14]

Quantitative trait loci mapping

Quantitative trait loci analysis was performed for each population separately using Map QTLversion 6.[15] Quantitative trait loci were detected using the interval mapping (IM) approach, mean values and the previously published reference genetic map generated from two populations.[16] Significance LOD score thresholds were calculated through permutation of 1000 iterations with an alpha risk of 0.05 and confidence limit of 95%. Confidence intervals of QTL positions were determined as two-LOD support intervals.

Quantitative trait loci validation

Based on the genotypic and phenotypic data from the diversity panel, QTL validation was performed in two steps. First, a simple linear model that associates each phenotype with a SNP was tested using the Pearson correlation test and a 5% significance level. All markers included in the QTL confidence intervals were tested. The Pearson coefficient of determination (R^2) and the *P*-value (Pearson) of each SNP marker were thus determined. Second, an ANOVA was performed for the significant markers to determine whether the observed genotypic classes were significantly different at $P < 0.05$.

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

Identification of candidate genes

Candidate genes for the validated QTLs were identified by searching near the significant SNPs in the NCBI database, which contains all 35 078 genes that were annotated on the *D. rotundata* reference genome.[17]

RESULTS

Phenotyping of mapping populations

The mean values and ranges of scores for starch, DMC, sugar, and protein content in the two mapping populations are presented in Table 1. Starch content ranged from 74.0% to 83.6% in population A (74F × Kabusa) and from 71.1% to 85.6% in population B (74F × 14 M). For DMC, the range of scores was from 26.7% to 38.8% in population A and from 27.6% to 41.3% in population B. Sugar content ranged from 0.58% to 5.08% in population A and from 0.10% to 5.84% in population B. Finally, protein content scores ranged from 3.95% to 7.64% and from 4.05% to 7.90% in populations A and B, respectively. In male progenitors, 14 M and Kabusa, the values were significantly different from those of female 74F for starch (80.5%, 79.2%, and 77.2%, respectively), DMC (31.5%, 28.9%, and 28.1%), and sugars (1.79%, 4.12%, and 3.22%).

The ANOVA on phenotypic data of population A showed highly significant ($P < 0.0001$) genotype effects and significant repetition effects ($P < 0.05$) for all traits (Table 1). However, genotypes were the most important source of variation for each studied trait. An ANOVA of phenotypic data from population B also showed highly significant genotypic effects ($P < 0.0001$) for all studied traits (Table 1).

The frequency distribution of all traits studied showed typical quantitative variation in both mapping populations and all traits fitted a normal distribution (Fig. 1). Transgressive segregations were observed with lower or higher phenotypic values than those of the parents for all traits (Fig. 1).

Broad heritability for the four traits ranged from 0.68 to 0.88 in population B and from 0.69 to 0.81 in population A (Table 1). The heritability for starch was similar in both populations (0.68 and 0.69) and also for protein (0.78 and 0.81). The heritability obtained for DMC and sugar content was significantly higher in population B (0.86 and 0.88) than in population A (0.75 and 0.72).

Several significant correlations were detected between traits (Table 2). Negative correlations were found in both biparental populations between starch content and protein content, and between starch content and sugar content. A positive correlation was detected in population A between starch content and DMC, while in population B, a positive correlation was detected between DMC and sugars.

Quantitative trait loci detection

A total of 25 QTLs were identified for the four studied traits in both mapping populations. For starch content, four QTLs were detected in the population B located on chromosomes 2, 5, 6, and 10, which explained 22.4, 22.6, 20.1, and 22.1% of phenotypic variance, respectively (Table 3). In population A, three QTLs were detected on chromosomes 10, 11, and 18, which explained 15.4, 19.1, and 16.1%, respectively of total phenotypic variance. Both QTLs identified on chromosome 10 were located in distinct regions and are two different loci. Figure 2(a) shows the one that was identified in population B.

For DMC, four QTLs were identified in population B located on chromosomes 1, 4, 7, and 12, which explained 98.5% of total phenotypic variance. In population A, two QTLs were found on chromosomes 1 and 2, explaining 14.3% and 18.7% of phenotypic variance, respectively. Figure 2 shows information from the QTL identified on chromosome 2 (population A).

Three QTLs were detected for sugars in population B (chromosomes 7, 9, and 13) and three in population A (chromosomes 6, 7 and 12), explaining 68.5% and 58.8% of total phenotypic variance in each population, respectively. Figure 2 shows results from the QTL found in chromosome 7 (population A).

Finally, four QTLs were revealed for proteins in population B (chromosomes 2, 5, 8, and 19) and two in population A (chromosomes 10 and 18), which explained 98.4% and 38.4% of each population's total phenotypic variance. Figure 2 shows the results from the QTL found on chromosome 19.

Quantitative trait loci validation

The majority of QTLs (22 of 25) were validated in the diversity panel. Table 3 presents the SNP markers located within confidence intervals of the QTLs, showing a significant association with the diversity panel phenotypic data (at $P < 0.01$** or $P < 0.05$*). The alleles at each locus and the allelic effects are also presented in Table 3. The analyses of variance showed that differences between the different genotypic classes were significant ($P < 0.05$) for all validated QTLs. Figure 2(b) presents the phenotypic data distributions of the different genotypic classes for QTLs

Table 1. Phenotypic variation for starch, DMC, sugar and protein content in mapping populations A (74F × Kabusa) and B (74F × 14 M)

Trait	Pop	Mean	Min.	Max.	G	R	H^{2a}
Starch	B	78.7	71.1	85.6	***	-	0.68
DMC		33.3	27.6	41.3	***	-	0.86
Sugars		3.03	0.10	5.84	***	-	0.88
Proteins		5.74	4.05	7.90	***	-	0.78
Starch	A	79.8	74.0	83.6	***	*	0.69
DMC		32.7	26.7	38.8	***	*	0.75
Sugars		236	0.58	5.09	***	**	0.72
Proteins		5.55	3.95	7.64	***	*	0.81

Note: ***Significant at $P < 0.0001$, ** $P < 00.1$, * $P < 00.5$ for the effects of genotype (G) and repetition (R) on the phenotypic variance estimated by ANOVA.
[a] Broad –sense heritability. Starch, proteins and sugars are expressed as a percentage of dry weight, and dry matter content is expressed as a percentage of fresh weight.

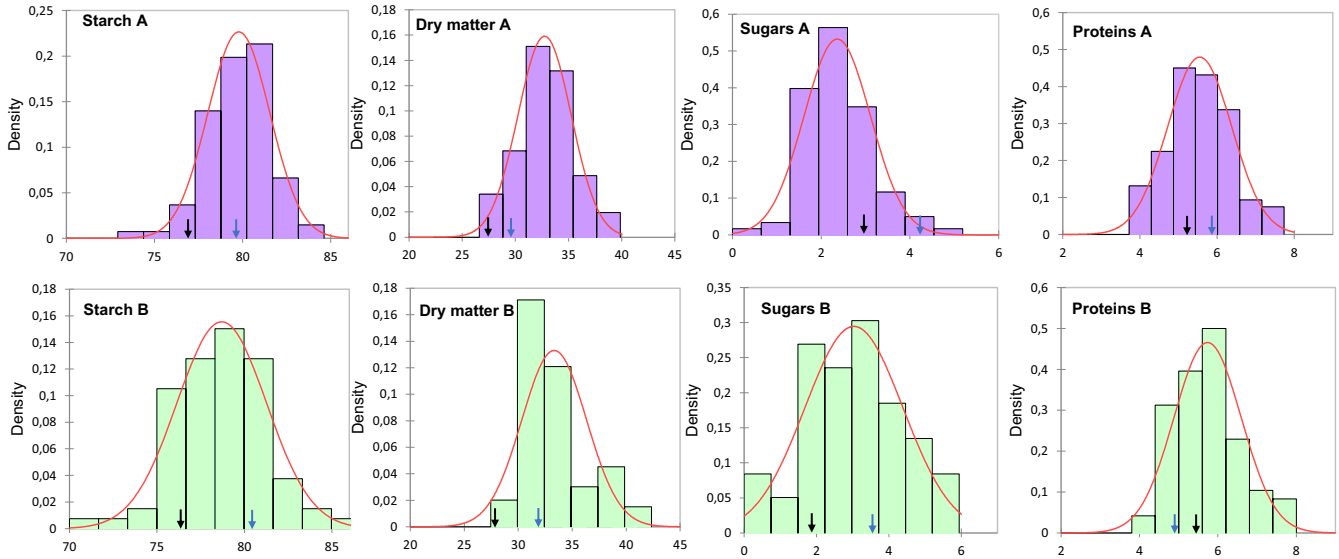


Figure 1. Phenotypic distributions for starch, dry matter content, sugar and protein content in mapping populations A (74F × Kabusa) and B (74F × 14 M).

Table 2. Coefficients of correlation between traits in biparental populations A (74F × Kabusa) and B (74F × 14 M)

Variable	Starch	Sugars	Proteins
Sugars	−0.34*		
	−0.57*		
Proteins	−0.66*	−0.11	
	−0.27*	−0.24	
DMC	0.22*	−0.15	−0.20
	−0.23	0.47*	0.00

Note: Top values are those from population A and bottom values from population B.
*Significant at $P < 0.01$.

detected for starch, proteins, sugars and DMC on chromosomes 10, 19, 7, and 2, respectively. These QTLs were used as examples to illustrate the kind of data obtained for the different traits. At locus 10.1.6610171, the allele C is associated with a lower starch content, whereas at locus 02.1.28100114, the allele A is associated with low dry matter content. At locus 21.1.307693 three different genotypic classes were observed and varieties homozygous for allele T (TT) had significantly higher protein content. At locus 07.1.1765193, varieties homozygous for allele T (TT) had significantly higher sugar content.

Identification of potential candidate genes

Candidate genes were detected for the majority of validated QTLs and they are presented in Table 3. The significant SNP markers were generally located in intergenic regions but four were found in intronic regions and one in an exon.

Five candidate genes were identified out of the seven QTLs for starch, including *Sucrose phosphatase 1* (EC 3.1.3.24), *Sucrose phosphatase synthase* (EC 2.4.1.14), *Isoamylase 3* (EC 3.2.1.68), *Glycosil hydrolase*, and *Serine/threonine protein kinase*.

Three candidate genes were identified out of the five QTLs for DMC, including a *Xyloglucan galactosyltransferase*, an *Ubiquitin specific protease* and an *Endo beta D Glucanase*.

A total of five candidate genes were identified out of six QTLs for sugars. They are *Pyruvate dehydrogenase kinase* (EC 1.2.4.1), *Enolase Chloroplastic* (EC 4.2.1.11), *beta glucosidase, rhamnose-galactose sugar transporter*, and *glycine cleavage system H protein*.

Finally, four candidate genes for proteins were detected out of the five QTLs (Table 3). They are a *D amino acid transaminase, Protein activity of BC1 complex kinase 7*, *Indole-3-Glycerol phosphate synthase* (EC 4.1.1.48), and *Proline rich receptor like protein kinase*.

DISCUSSION

This research provides valuable insights into the genetic architecture of key quality related traits in yam; identifies markers linked to the genomic regions controlling each trait, which is useful for MAS; validates the QTLs on a diversity panel; and identifies candidate genes from the validated QTLs.

A larger phenotypic variability was observed in population B (74F X 14 M) than in population A (74F X Kabusa) for DMC, starch, and sugars. This could be explained by greater differences observed between progenitors of this population. In both progenies the presence of transgressive hybrids (having higher or lower values than those of their parents) could be due to the heterozygosity of progenitors, and in particular of female 74F and male 14 M. Indeed, the high heterozygosity found for these two genotypes[16] would favor a high frequency of new allelic combinations, thus widening of phenotypic variation within their progenies. A similar phenomenon was observed in a previous study focusing on the same populations but on other tuber quality traits.[11]

Heritabilities of all traits were moderately high to high, with 68% to 88% of the phenotypic variation in hybrid means due to genetic differences between hybrids. This makes selection for these quality traits possible in breeding programs. The high heritability for DMC was similar to that found in 191 genotypes of different *Dioscorea* species (0.86).[12] As expected, starch content was the trait with the lowest heritabilitiy. Indeed, it is known that starch content varies between different parts of tuber ('head, middle, bottom')[18] and also during the storage of tubers.[19] Interestingly, starch was negatively correlated with proteins and sugars

Table 3. Quantitative trait loci validated for starch, DMC, sugar, and protein content in the diversity panel, and candidate genes identified

Trait	Chr	Pop.	LOD	R2 (%)	SNP	Alleles /Allele effect	Localization	Candidate gene
Starch	5	B	3.67	22.6	05.1_32707706**	CC/CT/TT (T-)	Intergenic	*G- type lectin Receptor like serine/threonine protein kinase*
Starch	2	B	3.63	22.4	02.1_32141722*	TG/GG (G-)		
Starch	6	B	3.39	20.1	06.1_20308422*	AA/A*/**(*+)	Intergenic	*Sucrose phosphatase 1 EC 3.1.3.24*
Starch	10	B	3.58	22.1	10.1_6610171*	T*/**(*-)	Intergenic	*Glycosil hydrolase family 5*
Starch	10	A	3.24	15.4	10.1_17600318*	CC/C* (*+)	Intergenic	*Sucrose phosphate synthase EC 2.4.1.14*
Starch	11	A	4.10	19.1	11.1_4493720**	TT/TC (C-)	Intronic	*Uncharacterized protein Loc120271825*
Starch	18	A	3.40	16.1	18.1_1695612*	GG/GA/AA (A+)	Intronic	*Isoamylase 3 chloroplastic EC 3.2.1.68*
Proteins	8	B	4.46	26.6	08.1_18931363*	CC/CG (G-)	Intergenic	*D amino acid transaminase EC2.6.1.21*
Proteins	19	B	4.33	26.1	21.1_107691*	CC/CT/TT (T+)	Intergenic	*Protein activity of BC1 complex kinase 7*
Proteins	5	B	3.85	23.6	05.1_31811203*	TT/T* (*-)	Intergenic	*Indole-3-Glycerol phosphate synthase EC 4.1.1.48*
Proteins	2	B	3.54	21.9	-	non val.		
Proteins	18	A	4.90	22.4	18.1_1629151**	GG/GA/AA (A+)	Exon	*Uncharacterized protein Loc120282302*
Proteins	10	A	3.36	16.1	10.1_16586803**	GG/GC/CC (C-)	Intergenic	*Putative proline rich receptor like protein kinase*
Sugars	9	B	4.07	24.7	09.1_19344114**	A*/** (*-)	Intergenic	*Enolase 1 chloroplastic EC 4.2.1.11*
Sugars	7	B	3.91	23.9	07.1_351152*	GG/GA/AA (A-)	Intronic	*Pyruvate dehydrogenase kinase EC 1.2.4.1*
Sugars	13	B	3.18	19.9	-	non val.		
Sugars	12	A	4.60	21.2	12.1_17583120**	TG/GG (G+)	Intergenic	*Alpha 1 Arabinofuranosidase 1 EC 3.2.1.55*
Sugars	6	A	4.49	20.7	06.1_18785797	AA/A* /** (*-)	Intergenic	*Glycine cleavage system H Protein 2*
Sugars	7	A	3.51	16.6	07.1_1765193**	CT/TT (T+)	Intergenic	*Beta glucosidase 22*
DMC	4	B	4.84	27.2	04.1_10157871*	TT/TC/CC (C-)	Intergenic	*Xyloglucan galactosyltransferase GT17*
DMC	12	B	4.77	25.1	12.1_24347534*	TT/TC/CC(C-)		
DMC	7	B	4.13	24.0	07.1_3805155*	GG/GA (A+)	Intergenic	*Endo 1,3 (4) beta D Glucanase*
DMC	1	B	3.79	22.2	01.1_601071*	GG/**/** (*-)	Intronic	*Ubiquitin like specific protease 2B*
DMC	1	A	3.00	14.3	-	non val.		
DMC	2	A	3.99	18.7	02.1_28100114**	GA/AA (A-)		

Note: *Chr* Chromosome, *Pop* Population, *SNP* marker in the QTL confidence interval that showed a significant association with phenotypic data from the diversity panel.
Note: **Significant at $P < 00.1$, * $P < 00.5$.
Note: *Non val* QTLs not validated in the diversity panel.

in both populations. This is consistent with previous results obtained on *D. alata*[9] and other root and tuber crops.[20]

The current research has led to the identification of 25 QTLs associated with the genetic variation of these four important tuber quality traits. Several QTL co-localizations were congruent with the observed genetic correlations. Three QTL co-localizations were observed for starch and proteins on chromosomes 18, 10, and 5. In addition, one QTL co-localization was observed for starch and sugars on chromosome 6. The starch QTL on chromosome 18 was located within 67 kb of the protein QTL. This short distance separating the QTLs suggests that these could be under the control of a same gene with a pleiotropic effect or several distinct, closely related genes. Starch QTLs on chromosomes 10 and 5 were at 1 Mb and 896 kb of protein QTLs, respectively. The QTL for starch on chromosome 6 was approximately 1.5 Mb from the sugar QTL. For these more distant QTLs, the hypothesis of a control by separate genes seems the most likely. Despite the existence of a negative correlation between starch and proteins, several hybrids containing both high starch and protein content were detected in both biparental populations. This can be explained by genetic recombinations between QTLs, which is feasible considering their physical distances and its telomeric chromosomal localization. Furthermore, no co-localization was observed for QTLs detected for a same trait in both populations on the same chromosome (starch on chromosome 10, sugars on chromosome 7, and DMC on chromosome 1), which is in accordance with the hypothesis that these are different loci.

Before being used for MAS, a QTL needs to be validated to confirm that its effect can be also detected in different genetic backgrounds. For this purpose, we used a contrasting diversity panel. A total of 22 QTLs could be validated showing that they are not specific to the genetic background of progenitors. These should be useful for breeding programs using MAS to select the favorable alleles and to improve yam tuber quality.

In addition, the validation process was very useful to identify the candidate genes, in particular when QTL confidence intervals were large. A total of five putative candidate genes were detected near the markers associated with starch content of which four genes (*Isoamylase ISA3, Serine/threonine protein kinase, Sucrose phosphate synthase, Sucrose phosphatase*) play important roles in the starch and sucrose metabolism. *Isoamylase ISA3* was reported to participate in the process of starch degradation in potatoes.[21] This gene was also reported to be involved in the reduction of dormancy period.[21] Both *Sucrose phosphate synthase* and *Sucrose phosphatase* play important roles in the sucrose metabolism.[22] The fourth putative gene, *Serine/threonine-protein kinase* was reported to participate in the process of starch and sugar biosynthesis in potato.[23] The fifth putative gene, *Glycosil hydrolase family*

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

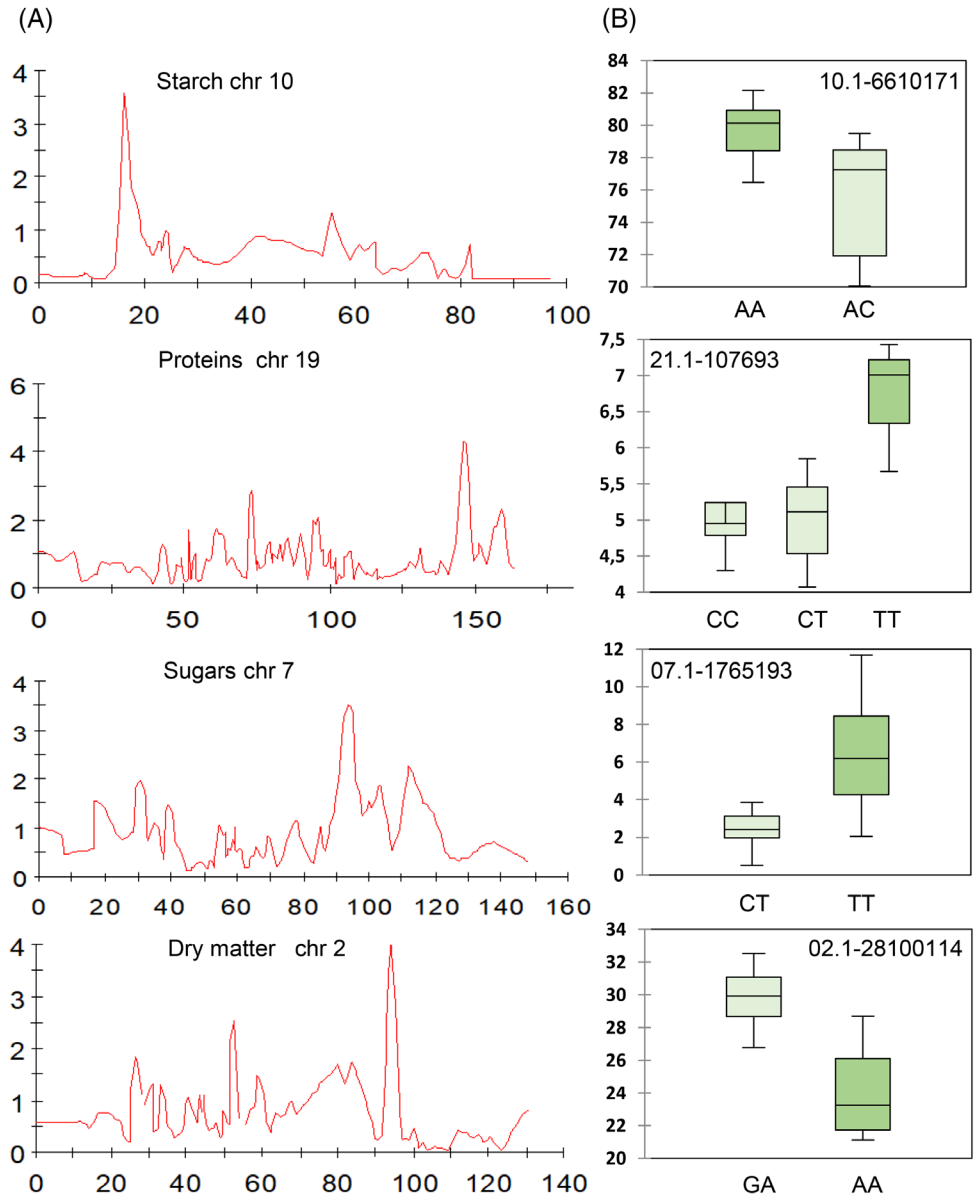


Figure 2. (A) Examples of QTLs detected in the biparental populations for starch, proteins, sugars and dry matter on chromosomes 10, 19, 7, and 2 respectively. (B) SNPs markers in the QTL confidence interval that showed a significant association with phenotypic data from diversity panel. The alleles at each locus and the allelic effects are depicted.

5, was reported to hydrolyze the glycosidic bond between two or more carbohydrates, and is present in many plant tissues.[24,25]

Five putative candidate genes were detected for sugar content, of which three play an important role in plant respiration or glycolysis (*Enolase, Pyruvate dehydrogenase kinase, Glycine H Protein*). *Pyruvate dehydrogenase kinase* was reported to be a negative regulator of the mitochondrial pyruvate dehydrogenase that plays a central role in control of cell respiration.[26] *Enolase*, also known as *Phosphopyruvate hydratase*, catalyzes the ninth and penultimate step of glycolysis. In tobacco, the glycine H protein was reported to play an important role in the photorespiratory flux.[27] Overexpression of this protein reduced the amounts of soluble sugars and increased the accumulation of starch.[27] The fourth putative gene *Beta Glucosidase* was reported to be involved in the hydrolysis of cellobiose and other oligosaccharides into glucose.[28] The fifth putative gene *UDP-rhamose/UDP-galactose transporter* was reported to be involved in the transport of nucleotide sugars (*UDP rhamose, UDP galactose*) from the cytosol into the Golgi lumen to be used in the synthesis of polysaccharides.[29]

Three putative candidate genes were identified for DMC. Of these three genes, *Xyloglucan galactosyltransferase* was reported to play a significant role in enhancing the plasticity of cell wall components through its ability to hydrolyze and reconnect the xyloglucan chains.[30] The second candidate gene, *endo Beta D Glucanase*, was reported to be involved in the cleavage of glucan chains, which are major constituents in cell walls, generating mainly oligosaccharides.[31] The third putative gene, *Ubiquitin protease*, was reported to play an important role in many plant developmental processes.[32]

Four putative candidate genes were detected for proteins. Of these four genes, *D-amino acid transaminase* was reported to play a crucial role in the biosynthesis and/or degradation metabolism

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

of different amino acids in plants, such as alanine and serine.[33] The second candidate gene, *Indole 3 glycerol phosphate synthase*, was reported to be involved in the tryptophan biosynthesis.[34] The third putative gene *Complex kinase 7* was reported to be involved in the phosphorylation of proteins both in the mitochondrial outer membrane and in chloroplasts.[35] The fourth putative gene, *Proline-rich like receptor kinase*, belongs to the hydroxyproline-rich glycoprotein (HRGP) superfamily, which was reported to be involved in many plant developmental processes.[36]

The candidate genes identified should be useful in providing a better understanding of the physiological and molecular basis of these important tuber quality traits.

AUTHOR CONTRIBUTIONS

Arnau G designed the study with support from Chair H; Ehounou AE, Maledon E, Nudol E and Leinster J contributed to sample preparation for NIRS analysis. Arnau G, Ehounou AE, Maledon E, and Nudol E contributed to sample preparation for chemical analysis. Desfontaines L and Marie-Magdeleine C generated the NIRS spectroscopic data. Arnau G performed QTL mapping, validation analysis, and candidate gene research. All authors contributed to the writing of the manuscript.

ACKNOWLEDGEMENTS

This work was supported by the RTBfoods project https://rtbfoods.cirad.fr, through a grant, OPP1178942: Breeding RTB products for end user preferences (RTBfoods) to the French Agricultural Research Centre for International Development (CIRAD), Montpellier, France, by the Bill and Melinda GatesFoundation (BMGF). A special thanks to Hernan Cerballos for his helpful comments and improvement of the manuscript and to Clair Hershey for English proofing.

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1 FAO statistical database, https://www.fao.org (2022).
2 Cornet D, Sierra J, Tournebize R, Gabrielle B and Lewis FI, Bayesian network modeling of early growth stages explains yam interplant yield variability and allows for agronomic improvements in West Africa. *Eur J Agron* **75**:80–88 (2016).
3 Baah FD, Maziya-Dixon B, Asiedu R, Oduro I and Ellis WO, Nutritional and biochemical composition of D. alata (*Dioscorea* spp.) tubers. *J Food Agric Environ* **7**:373–378 (2019).
4 Egesi CN, Asiedu R, Egunjobi JK and Bokanga M, Genetic diversity of organoleptic properties in water yam (*Dioscorea alata* L). *J Sci Food Agric* **83**:858–865 (2003).
5 Otegbayo B, Tanimola A, Oroniran O, Maraval I, Forestier-Chiron N and Bugaud C, Sensory Characterization of Pounded Yam. Biophysical Characterization of Quality Traits, WP2. Iwo, Nigeria, RTBfoods Laboratory standard operating procedure, 15p (2020). https://doi.org/10.18167/agritrop/0059.
6 Adinsi L, Akissoe N, Maraval I, Forestier-Chiron N and Bugaud C, Sensory Characterization of Boiled Yam. Biophysical Characterization of Quality Traits, WP2. Cotonou, Benin,RTBfoods Laboratory standard operating procedure, 13p (2020). https://doi.org/10.18167/agritrop/00600.
7 Lebot V, Malapa R, Molisalé T and Marchand JL, Physico-chemical characterisation of yam (*Dioscorea alata* L.) tubers from Vanuatu. *Genet Resour Crop Evol* **53**:1199–1208 (2006).
8 Lebot V, Lawac F and Legendre L, The greater yam (*Dioscorea alata* L.): a review of its phytochemical content and potential for processed products and biofortification. *J Food Compos Anal* **115**:104987 (2023).
9 Lebot V and Malapa R, Application of near infrared reflectance spectroscopy for the evaluation of yam (*Dioscorea alata*) germplasm and breeding lines: quality evaluation of yam using NIRS. *J Sci Food Agric* **93**:1788–1797 (2013).
10 Ehounou AE, Cornet D, Desfontaines L, Marie-Magdeleine C, Maledon E, Nudol E *et al.*, Predicting quality, texture and chemical content of yam (*Dioscorea alata* L.) tubers using near infrared spectroscopy. *J Near Infrared Spectrosc* **29**:128–139 (2021).
11 Ehounou AE, Cormier F, Maledon E, Nudol E, Vignes H, Gravillon MC *et al.*, Identification and validation of QTLs for tuber quality related traits in greater yam (*Dioscorea alata* L.). *Sci Rep* **12**:8423 (2022). https://doi.org/10.1038/s41598-022-12135-2.
12 Adejumobi II, Agre PA, Onautshu DO, Adheka JG, Cipriano IM, Monzenga JC *et al.*, Assessment of the yam landraces (*Dioscorea* spp.) of DR Congo for reactions to pathological diseases, yield potential, and tuber quality characteristics. *Agriculture* **12**:599 (2022). https://doi.org/10.3390/agriculture12050599.
13 Gatarira C, Agre P, Matsumoto R, Edemodu A, Adetimirin V, Bhattacharjee R *et al.*, Genome-wide association analysis for tuber dry matter and oxidative browning in water yam (*Dioscorea alata* L.). *Plants (Basel)* **9**:969 (2020). https://doi.org/10.3390/plants9080969.
14 Sharif BM, Burgarella C, Cormier F, Fabien C, Mournet P, Causse S *et al.*, Genome-wide genotyping elucidates the geographical diversification and dispersal of the polyploid and clonally propagated yam (*Dioscorea alata* L.). *Ann Bot* **6**:1029–1038 (2020).
15 Van Ooijen JW, *MapQTL 6, Software for the Mapping of Quantitative Trait Loci in Experimental Populations of Diploid*. Wageningen, Netherlands, Kyazma BV (2009).
16 Cormier F, Lawac F, Maledon E, Gravillon MC, Nudol E, Mournet P *et al.*, A reference high-density genetic map of greater yam (Dioscorea alata L.). *Theor Appl Genet* **132**:1733–1744 (2019).
17 Tamiru M, Natsume S, Takagi H, White B, Yaegashi H, Shimizu M *et al.*, Genome sequencing of the staple food crop white Guinea yam enables the development of a molecular markerfor sex determination. *BMC Biol* **15**:86 (2017). https://doi.org/10.1186/s12915-017-0419-x.
18 Afoakwa EO, Polycarp D, Budu AS, Mensah-Brown H and Otoo E, Variability in biochemical composition and cell wall constituents among seven varieties in Ghanaian yam (Dioscorea sp.) germplasm. *Afr J Food Agric Nutr Dev* **13**:8106–8127 (2013). https://doi.org/10.18697/ajfand.59.13280.
19 Wireko-Manu FD, Oduro I, Ellis WO, Asiedu R and Maziya-Dixon B, Food quality changes in water yam (Dioscorea Alata) during growth and storage. *Asian J Agric Food Sci* **1**:66–72 (2013).
20 Lebot V, Malapa R and Jung M, Use of NIRS for the rapid prediction of total N, minerals, sugars and starch in tropical root and tuber crops. *N Z J Crop Hortic Sci* **41**:144–153 (2013). https://doi.org/10.1080/01140671.2013.798335.
21 Ferreira SJ, Senning M, Fischer-Stettler M, Streb S, Ast M, Neuhaus HE *et al.*, Simultaneous silencing ofisoamylases ISA1, ISA2 and ISA3 by multi-target RNAi in potato tubers leads to decreased starch content and an early sprouting phenotype. *PLoS One* **12**:e0181444 (2017). https://doi.org/10.1371/journal.pone.0181444.
22 Fernie AR, Willmitzer L and Trethewey RN, Sucrose to starch: a transition in molecular plant physiology. *Trends Plant Sci* **7**:35–41 (2002). https://doi.org/10.1016/s1360-1385(01)02183-5 PMID: 11804825.
23 McKibbin RS, Muttucumaru N, Paul MJ, Powers SJ *et al.*, Production of high-starch, low-glucose potatoes through over-expression of the metabolic regulator SnRK1. *Plant Biotechnol J* **4**:409–418 (2006).
24 Neis A and da Silva Pinto L, Glycosyl hydrolases family 5, subfamily 5: relevance and structural insights for designing improved biomass degrading cocktails. *Int J Biol Macromol* **193**:980–995 (2021). ISSN 0141-8130). https://doi.org/10.1016/j.ijbiomac.2021.10.062.
25 Minic Z, Physiological roles of plant glycoside hydrolases. *Planta* **227**:723–740 (2008).
26 Millar AH, Knorpp C, Leaver CJ, Christopher JC and Hill SA, Plant mitochondrial pyruvate dehydrogenase complex: purification and

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

identification of catalytic components in potato. *Biochem J* **334**:571–576 (1998).
27 López-Calcagno PE, Fisk S, Brown KL, Bull SE, South PF and Raines CA, Overexpressing the H-protein of the glycine cleavage system increases biomass yield in glasshouse and field-grown transgenic tobacco plants. *Plant Biotechnol J* **17**:141–151 (2019). https://doi.org/10.1111/pbi.12953 Epub 2018 Jul 22. PMID: 29851213; PMCID: PMC6330538.
28 Angkasawati VS, Kamaliyah I, Putra MD, Nata IF and Irawan C, Structural characterization and antioxidant activity of liquid sugar from Alabio potato using enzymatic hydrolysis processes. *IOP Conf Ser: Mater Sci Eng* **980**:012041 (2020). https://doi.org/10.1088/1757-899X/980/1/0120412.
29 Rautengartena C, Ebert B, Moreno I, Templec H, Hertera T, Linke B *et al.*, The Golgi UDP-Rhamnose / UDP-Galactose transporter family in Arabidopsis. *Proc Natl Acad Sci U S A* **111**:11563–11568 (2014). https://doi.org/10.1073/pnas.1406073111.
30 Baumann MJ, Eklöf JM, Michel G, Kallas AM, Teeri TT, Czjzek M *et al.*, Structural evidence for the evolution of Xyloglucanase activity from xyloglucan Endo-Transglycosylases: biological implications for cell wall metabolism. *Plant Cell* **19**:1947–1963 (2007). https://doi.org/10.1105/tpc.107.051391.
31 Perrot T, Pauly M and Ramírez V, Emerging roles of -Glucanases in plant development and adaptative responses. *Plants* **11**:1119 (2022). https://doi.org/10.3390/plants11091119.
32 Zhou H, Zhao J, Cai J and Patil SB, Ubiquitin-specific proteases function in plant development and stress responses. *Plant Mol Biol* **94**:565–576 (2017). https://doi.org/10.1007/s11103-017-0633-5.
33 Diebold R, Schuster J, Däschner JK and Binder S, The branched-chain amino acid transaminase gene family in Arabidopsis encodes plastid and mitochondrial proteins. *Plant Physiol* **129**:540–550 (2002).
34 Schlee S, Dietrich S, Kurćon T, Delaney P, Goodey NM and Sterner R, Kinetic mechanism of Indole-3-glycerol phosphate synthas. *Biochemistry* **52**:132–142 (2013).
35 Pagliarin DJ and Dixon JE, Mitochondrial modulation: reversible phosphorylation takes center stage. *Trends Biochem Sci* **31**:26–34 (2006) ISSN 0968–0004). https://doi.org/10.1016/j.tibs.2005.11.005.
36 Borassi C, Sede AR, Mecchia MA, Mangano S *et al.*, Proline-rich extensin-like receptor kinases PERK5 and PERK12 are involved in pollen tube growth. *FEBS Lett* **595**:2593–2607 (2021). https://doi.org/10.1002/1873-3468.14185.

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.