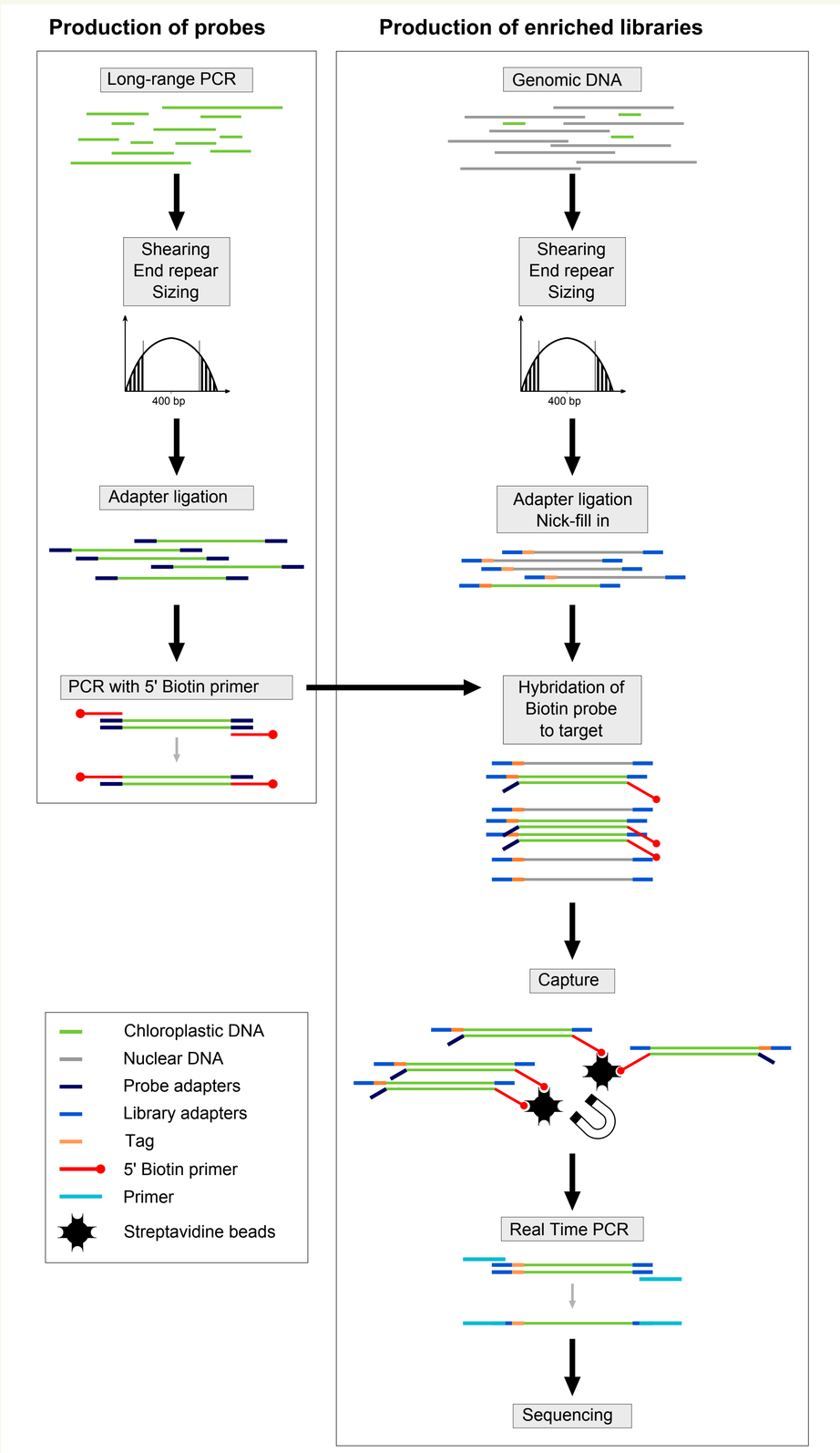
Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies

Cédric Mariac[1], Marie Couderc[1], Nora Scarcelli[1], Claire Billot[2], Adeline Barnaud[3], Adama Faye[4,5], Bonaventure Sonké[5], Jean-Christophe Pintaud[1], Yves Vigouroux[1] and Thomas L. P. Couvreur[1,5]
(1) IRD, Montpellier, France; (2) CIRAD, Montpellier, France; (3) IRD/ISRA, Dakar, Senegal; (4) IRD Yaoundé, Cameroon; (5) Ecole Normale Supérieure, University of Yaoundé I, Cameroon

Biodiversity, phylogeography and population genetics studies are expected to be revolutionized by access to large datesets thanks to new sequencing methods. We develop an easy and cost-effective protocol for in-solution enrichment hybridization capture of complete chloroplast genomes. The protocol uses cheap in-house species-specific probes developed via long range PCR of the entire chloroplast. Barcoded libraries are constructed and in-solution enrichment of the chloroplasts is done using the probe. Our protocol allows whole chloroplast genomes to be sequenced at a modest cost for large samples. This will allow unprecedented resolution for closely related species in phylogeography and population genetics studies using chloroplast sequences.

Protocol

Protocol undertaken on four crop species: *Oryza glaberrima* (African rice), *Pennisetum glaucum* (Pearl millet), *Dioscorea* ssp. (African yams) and *Digitaria exilis* (Fonio).
To capture chloroplast sequences directly from the genomic DNA libraries, biotinylated probes were produced in house.
Libraries were then constructed using 6-bp barcodes to allow for multiplexing at different degrees.



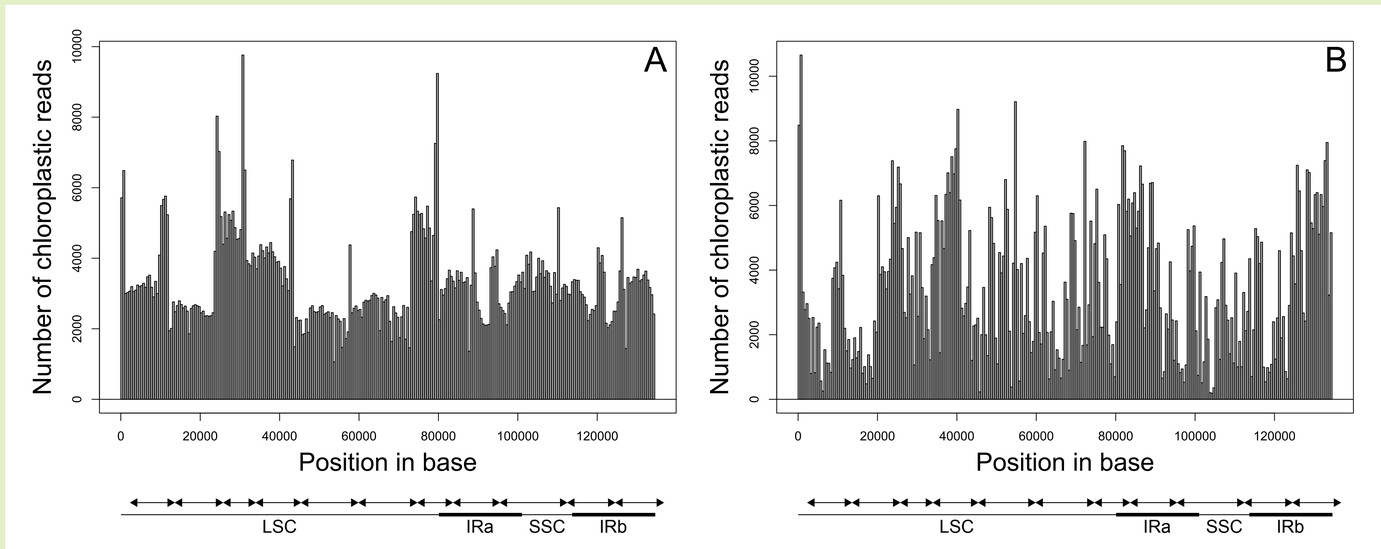
Probes and libraries construction for MiSeq Illumina sequencing

Conclusions

Hybridization via probes is a cost effective way to enrich and sequence entire organelle genomes at **deep multiplexing levels**.
Cost for a plastome (with library preparation): 22$ / individual.
Sequencing **95 plastomes in a single MiSeq run is achievable using our protocol**.
The protocol developed here allows the **analysis of plastid diversity** within and between closely related species in a cost and time effective way.

Results

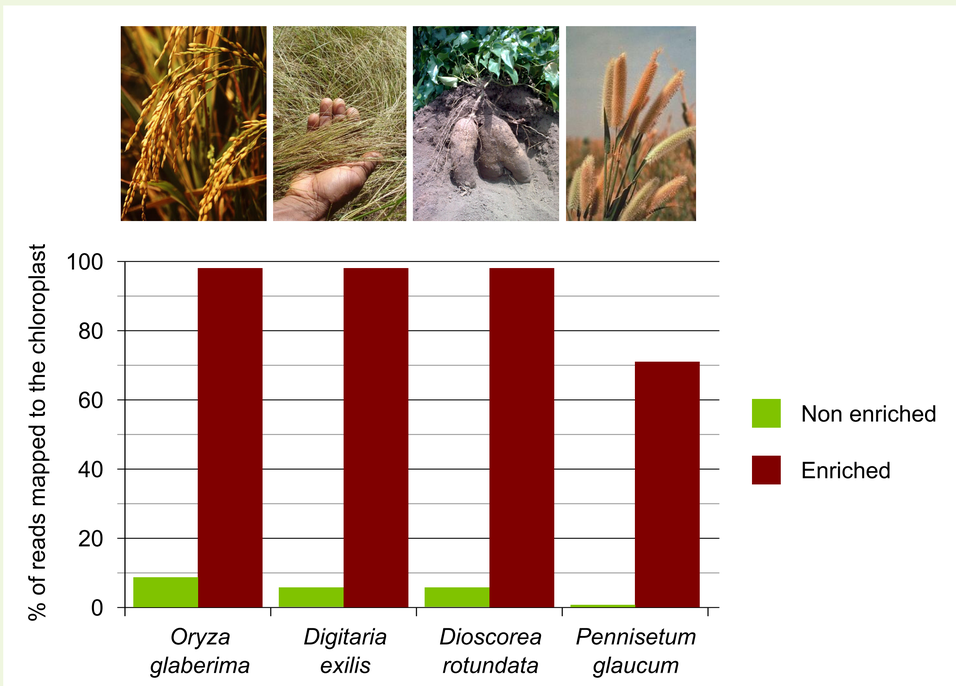
Validation of the probes in *Oryza glaberrima*



Number of reads mapped to the chloroplastic reference genome of *O. sativa* (NC_001320):
(A) Sequencing the probes only: confirms the probe is chloroplastic.
(B) Sequencing a library enriched using the probes: confirms good sequencing of chloroplast.

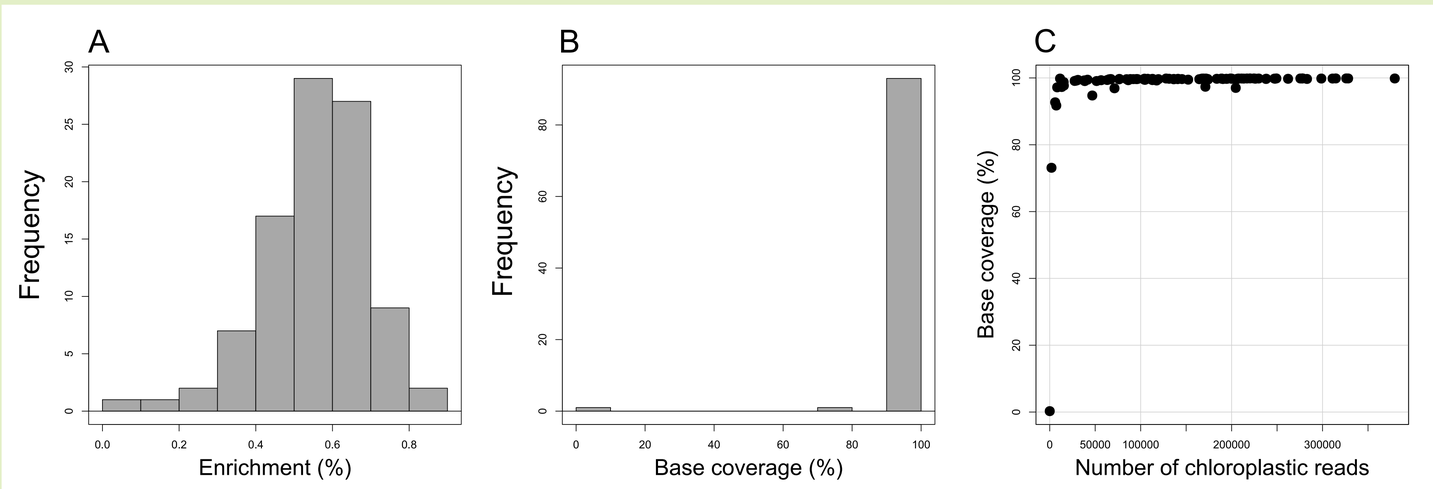
Position of long range PCR products are illustrated with arrowed horizontal lines. The position of Large single copy (LSC), inverted repeat (thick lines, IRa and IRb) and small single copy regions (SSC) are also reported

Enrichment success for 4 crop species



Enriched versus non enriched libraries for the four species using the enrichement protocol. Specific probes were designed for each species.

Enrichment success of 95 multiplexed pearl millet individuals in a single MiSeq run



(A) % of reads mapping to the reference; (B) Average coverage: 99% across the 94 individuals. 90 out of 95 with coverage > 95% ; (C) The number of reads mapping to their reference plotted against % cp genome coverage.

This study was financed by agropolis fondation grant number CHLORODIV 1202-040