

The South Green Bioinformatics platform, a comprehensive resource for crop genomics

Manuel Ruiz ¹, Alexis Dereeper ², Gaëtan Droc ¹, Gautier Sarah ³, Christine Tranchant-Dubreuil ⁴, Jean-François Dufayard ¹, Valentin Guignon ⁵, Chantal Hamelin ¹, Frédéric de Lamotte ³, Pierre Larmande ⁴, François Sabot ⁴, Guilhem Sempéré ⁶, Marilyne Summo ¹, Bertrand Pitollat ¹, Dominique This ⁸, Stéphanie Bocs ¹, Yann Hueber ⁵, Nordine El Hassouni ³, Guillaume Martin ¹, Anestis Gkanogiannis ¹, Sébastien Ravel ⁷, Cécile Monat ⁴, Ndomassi Tando ⁴, Mathieu Rouard ⁵

¹ CIRAD, UMR AGAP, Montpellier, France, ² IRD, UMR IPME, Montpellier, France, ³ INRA, UMR AGAP, Montpellier, France, ⁴ IRD, UMR DIADE, Montpellier, France, ⁵ Bioversity International, Montpellier, France, ⁶ CIRAD, UMR Intertryp, Montpellier, France, ⁷ CIRAD, UMR BGPI, Montpellier, France, ⁸ Montpellier SupAgro, UMR AGAP, Montpellier, France

Analysis and visualization of massive genomics datasets are an ongoing trend in plant sciences. The South Green Bioinformatics platform provides an ecosystem of tools that were originally developed as independent entities to fulfill the need for specific projects or crops, but have evolved over time to generic tools to comprehensively study crop genomics.

We have built a large panel of public information systems dedicated to specialized datasets (markers, genes, gene families, transcriptomes, genotypes, phenotypes, etc.) and crop-specific resources called Genome Hubs. Target users of bioinformatic analytical workflows are usually divided between people who use command-line and those who do not. We addressed both categories by offering complementary solutions, like Galaxy-based and command-line applications.

Various groups used the South Green infrastructure to obtain their data and results, and were able to publish high-quality biological information, on Coffee genome, Banana, Cocoa, African rice or large transcriptome resources. Tools developed for these studies are adaptable to a wide range of other organisms.