**Diversity and structure assessment of the genetic resources in a germplasm collection from a vanilla breeding program in Madagascar**

Rivo Onisoa Léa Rasoamanalina a\*, Khaled Mirzaei a, El Jaziri Mondher b Angel Rafael Ramirez Ramirez a,c, Pierre Bertin a

aEarth and Life Institute - Agronomy - Université catholique de Louvain - Belgium

bLaboratory of Plant Biotechnology, Université libre de Bruxelles, Belgium

cFaculty of Agroforestry, University of Guantánamo, Cuba

\*Corresponding author: [onyhrapanarivo@gmail.com](mailto:onyhrapanarivo@gmail.com)

Phone number: +26134 20 578 48

Croix du Sud 2 bte L7.05.11 -B-1348 Louvain-la-Neuve, Belgium

Bâtiment Kellner – second floor - D.291

**Table S1** Identifiers of the studied accessions, sequencing and mapping statistics of raw reads.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Accession ID** | **Name of taxon /Phenotype** | **N° BioSample accession** | **Number of sequenced**  **Reads** | **Aligned reads (%)** |
| sample\_02 | Tsitaitra | SAMN36015065 | 12540814 | 99.64 |
| sample\_03 | *V. pompona* | SAMN36015066 | 12965588 | 94.07 |
| sample\_04 | *V. planifolia* | SAMN36015067 | 23138364 | 99.64 |
| sample\_05 | Tsitaitra | SAMN36015068 | 1630588 | 99.6 |
| sample\_08 | *V. planifolia* | SAMN36015069 | 1529642 | 99.53 |
| sample\_09 | Tsitaitra | SAMN36015070 | 9852634 | 99.56 |
| sample\_10 | Tsitaitra | SAMN36015071 | 6086230 | 99.36 |
| sample\_11 | *V. planifolia* | SAMN36015072 | 10253026 | 99.54 |
| sample\_12 | *V. pompona* | SAMN36015073 | 4188718 | 99.37 |
| sample\_13 | *V. pompona* | SAMN36015074 | 9684316 | 99.67 |
| sample\_14 | Tsitaitra | SAMN36015075 | 102771728 | 99.64 |
| sample\_15 | Tsitaitra | SAMN36015076 | 50683744 | 99.56 |
| sample\_16 | Tsitaitra | SAMN36015077 | 88227150 | 99.58 |
| sample\_17 | *V. planifolia* | SAMN36015078 | 7078142 | 99.32 |
| sample\_19 | Vanille Banane | SAMN36015079 | 3063312 | 98.88 |
| sample\_20 | *V. pompona* | SAMN36015080 | 6966266 | 98.59 |
| sample\_21 | *V. planifolia* | SAMN36015081 | 13344648 | 99.68 |
| sample\_22 | *V. planifolia* | SAMN36015082 | 70053144 | 99.8 |
| sample\_23 | Tsitaitra | SAMN36015083 | 4337014 | 99.66 |
| sample\_24 | Tsitaitra | SAMN36015084 | 9481576 | 99.64 |
| sample\_27 | Vanille Banane | SAMN36015085 | 3791826 | 99.35 |
| sample\_29 | Tsitaitra | SAMN36015086 | 7856648 | 99.29 |
| sample\_30 | Tsivaky | SAMN36015087 | 17005890 | 99.67 |
| sample\_32 | Tsivaky | SAMN36015088 | 2332450 | 99.85 |
| sample\_33 | *V. planifolia* | SAMN36015089 | 40599960 | 99.65 |
| sample\_34 | *V. planifolia* | SAMN36015090 | 41642740 | 99.71 |
| sample\_35 | Tsitaitra | SAMN36015091 | 12139672 | 99.66 |
| sample\_36 | *V. pompona* | SAMN36015092 | 9924026 | 99.65 |
| sample\_37 | *V. pompona* | SAMN36015093 | 26560306 | 99.52 |
| sample\_38 | Tsitaitra | SAMN36015094 | 19364546 | 99.6 |
| sample\_39 | Tsitaitra | SAMN36015095 | 2774930 | 99.54 |
| sample\_40 | Tsitaitra | SAMN36015096 | 20456948 | 98.91 |
| sample\_41 | Tsitaitra | SAMN36015097 | 2147608 | 99.26 |
| sample\_42 | Tsitaitra | SAMN36015098 | 15731580 | 99.3 |
| sample\_43 | Tsitaitra | SAMN36015099 | 7257086 | 99.71 |
| sample\_44 | Tsitaitra | SAMN36015100 | 6061010 | 99.77 |
| sample\_45 | Tsitaitra | SAMN36015101 | 87024008 | 99.73 |
| sample\_46 | *V. planifolia* | SAMN36015102 | 8701124 | 99.6 |
| sample\_47 | *V. pompona* | SAMN36015103 | 5724580 | 99.59 |
| sample\_49 | Tsitaitra | SAMN36015104 | 1402164 | 99.74 |
| sample\_50 | *V. pompona* | SAMN36015105 | 21653682 | 99.48 |
| sample\_51 | Tsitaitra | SAMN36015106 | 10443220 | 99.49 |
| sample\_52 | Tsivaky | SAMN36015107 | 9510842 | 94.07 |
| sample\_53 | *V. planifolia* | SAMN36015108 | 2315862 | 99.64 |
| sample\_56 | Tsitaitra | SAMN36015109 | 2835012 | 99.6 |

**Table S2** Genomic location of the 23 701 common SNPs detected by Stacks, Bcftools, and GATK.

|  |  |
| --- | --- |
| **Consequence type** | **Count** |
| Intron SNP | 11 039 |
| Intergenic SNPs | 5 077 |
| Upstream gene SNPs | 2 627 |
| Downstream gene SNPs | 1 934 |
| Missense SNPs | 1 444 |
| Synonymous SNPs | 1 385 |
| Splice region SNPs | 169 |
| Stop gained SNPs | 16 |
| Stop lost SNPs | 3 |
| Splice donor SNPs | 3 |
| Spice acceptor SNPs | 1 |
| Start lost | 1 |

**Table S3** Comparison of the three variant callers: Stacks, BCFtools, and GATK. Time CPU: running time for variant calling (hours); Nb: number of filtered SNPs; He: expected heterozygosity; Ho: observed heterozygosity.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Pipeline** | **Time**  **CPU** | **Nb of filtered SNPs** | **% missing data (SD)** | **Loci coverage:**  **Mean (SD)** | **He** | **Ho** |
| Stacks | 2 | 316 398 | 17 (2.5) | 33.37 (17.8) | 0.12 | 0.14 |
| BCFtools | 39 | 80 888 | 14 (5.3) | 64.6 (23.4) | 0.24 | 0.34 |
| GATK | 316 | 126 391 | 13 (5.7) | 51.4 (15.6) | 0.23 | 0.30 |