Victoria Sichel[1], Gautier Sarah[1], Nabil Girollet[2], Valérie Laucou[1], Catherine Roux[1], Loïc Le Cunff[3], Patrice This[1]*, Pierre-François Bert[2], Thierry Lacombe[1, 4]

High quality phased assembly of grape genome offer new opportunities in chimera detection

Affiliations

[1] UMR AGAP Institut, Univ. Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France
[2] EGFV, Univ. Bordeaux, Bordeaux Sciences Agro, INRAE, ISVV, Villenave d'Ornon, France
[3] Institut Français de la Vigne et du Vin, Montpellier, France
[4] UMT Geno-Vigne®, IFV-INRAE-Institut Agro, Montpellier, France

Correspondence

Victoria Sichel: victoria.sichel@inrae.fr, Gautier Sarah: gautier.sarah@inrae.fr, Nabil Girollet: nabil.girollet@inrae.fr, Valérie Laucou: valerie.laucou@inrae.fr, Catherine Roux: catherine.roux@inrae.fr, Loïc Le Cunff: loic.lecunff@vignevin.com, Patrice This*: patrice.this@inrae.fr, Pierre-François Bert: pierre-francois.bert@inrae.fr, Thierry Lacombe: thierry.lacombe@supagro.fr

Summary

In perennial plants and especially those propagated through cuttings, several genotypes can coexist in a single individual, thus leading to chimeras. When the variant induces a noticeable phenotype modification, it can lead to a new cultivar. Viticulture already took economic advantage of this natural phenomenon: for instance, the berry skin of 'Pinot gris' derived from 'Pinot noir' by the selection of a chimera. Chimeras could also impact other crucial traits without being visually identified. Periclinal chimera where the variant has entirely colonized a cell layer is the most stable and can be propagated through cuttings. In grapevine, two functional cell layers are present in leaves, L1 and L2. However, lateral roots are formed from the L2 cell layer only. Thus, comparing DNA sequences of roots and leaves could allow chimera detection. In this study we used new generation Hifi long reads sequencing and recent bioinformatics tools applied to 'Merlot' to detect periclinal chimeras. Sequencing of 'Magdeleine Noire des Charentes' and 'Cabernet franc', the parents of 'Merlot', allowed haplotype resolved assembly. Pseudomolecules were built with few contigs, in some occasions only one per chromosome. This high resolution allowed haplotype comparison. Annotation from PN40024 was transferred to all pseudomolecules. Through variant detection, periclinal chimeras were found on both haplotypes. These results open new perspectives on chimera detection, which is an important resource to improve cultivars through clonal selection or breed new ones. Detailed results will be presented and discussed.

Key words

chimera, Hifi Sequencing, trio-binning, phased assembly, Whole genome

Introduction

A chimera is found when several sets of genotypes coexist in the same individual. The phenomenon was discovered in plants many centuries ago and has been fascinating scientist since then. They were used to understand plant ontogenesis (Satina, 1945, Szymkowiak and Sussex, 1996, Thompson and Olmo, 1963) but they are also an opportunity to breed new cultivars. In fact, viticulture has already been taking advantage out of chimeras by breeding different mutants of 'Pinot noir' under different cultivar names as the phenotype was so different. For instance, 'Pinot blanc' and 'Pinot gris' arose from two independent chimeras (Vezzulli *et al.*, 2012). Or again, the dwarf grapevine commonly used in grape research programs and was obtained performing somatic embryogenesis from 'Pinot Meunier' (Torregrosa *et al.*, 2019).

Chimeras are formed when a somatic genetic variation appears in a cell and is propagated through cell divisions. It can colonise an entire cell layer and is then called periclinal (Blakeslee *et al.*, 1939). In some rare cases, the mutation affects the phenotype and become visible (Bauhin, 1598, Nati, 1644).

Chimeras can spread all along the cell layer but can't jump from one cell layer to another unless there is a genetic accident. Therefore, by comparing two cell layers we can detect chimeras. Fortunately, grapevine leaves have two functional cell layers L1 and L2 whereas lateral roots come from a dedifferentiation of pericycle cells and only have the L2 cell layer (Pratt *et al.*, 1959, Thompson and Olmo, 1963). To compare L1 and L2 we need to sequence both roots and leaves from a non-grafted individual. But cell layers aren't the only challenge.

Up to now, chimera detection or confirmation has been possible when the mutation appeared on a heterozygous positions meaning that three alleles were found on a locus. On homozygous positions, chimeras could hardly be confirmed

Submitted/accepted for publication: March 30, 2023/September 20, 2023

because it could be confused with haplotype difference. Then the second challenge is obtaining a very high quality reference sequence and a haplotype phased assembly. In our case, we studied 'Merlot', which is a cross between 'Cabernet franc' and 'Magdeleine Noir des Charentes '(Boursiquot *et al.*, 2009).

Material and Methods

Leaf samples of 'Magdeleine Noire des Charentes' were collected from INRAE Vassal Montpellier (Marseillan, France), while 'Cabernet franc' leaves and 'Merlot' clone 343 leaves and roots came from IFV collection, Domaine de l'Espiguette (Le Grau-du-Roi, France). Long fragment DNA extraction was performed followed by PACBIO® hiFI long reads sequencing using Sequel II technology. 'Merlot' reads were sorted per haplotype by trio-binning. We then obtained two files for each 'Merlot' sample: "Merlot-hap-CF" and "Merlot-hap-MG". The reads were then assembled in contigs with hifiasm (Cheng *et al.*, 2021) and finally chromosomes were built on following PNv4 (Rustenholz *et al.*, 2021) model, which was completed by crossing contigs from the other haplotype. In order to have a functional approach, we transferred annotations from PN.vcost V3 with Liftoff software (Shumate, 2021). We also chose to work on the most stable part of the genome and took out all repeated sequences using Repeat masker (Smit *et al.*, 2013-2015). All reads were mapped on both 'Merlot' assemblies and variants were called using DeepVariant (Ryan Poplin, 2018). Chimera detection was performed by filtering the Variant Calling File (VCF) with vcftools program (Danecek *et al.*, 2011).

Periclinal L1 cell layer chimeras were selected when the variant was carried by only one haplotype on the leaf sample but all the other packets of reads had the reference nucleotide (Fig. 1). Mutation in the L2 cell layer had the variant on all reads from roots and a few reads from leaves, parental reads and opposite haplotype will have the same nucleotide as leaves but different from roots (Fig. 2).

Results

High quality reference genome

This results in a very high quality reference sequence per haplotype, where in some cases there is only one contig per chromosome and Liftoff correctly transferred 96% genes from PN.vcost V3 (Table 1).

Detected chimeras

When applying very selective criteria on SNV's outside repeated sequence, 104 positions were identified as chimeric throughout the entire genome, 33% were located in a gene body region and 15% in a coding region (Table 2).

According to the transferred annotation it was possible to see if the chimeras located in coding regions induced a modification in the corresponding amino acid. Out of the 16 positions, 11 changed the amino acid which could have a very important impact on the final protein conformation.

Discussion

Finally, this chimera detection method can be used both ways: from DNA sequencing to phenotype in order to identify non-visible chimeras or from phenotype to DNA sequencing to identify genetic regions controlling specific traits. Overall, chimeras are opportunities to understand metabolic pathways. They can lead to new cultivars if interesting agronomical characteristics are modified, and they also contribute

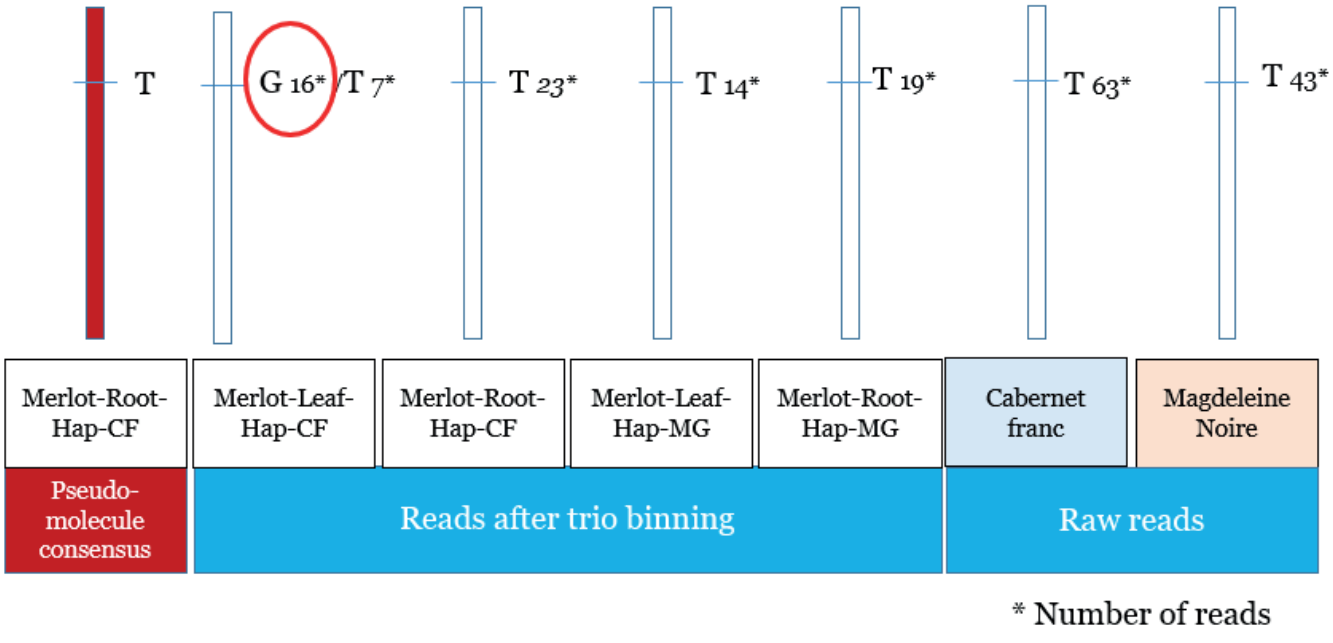

Fig. 1: An example of read distribution when a L1 chimeras was detected.

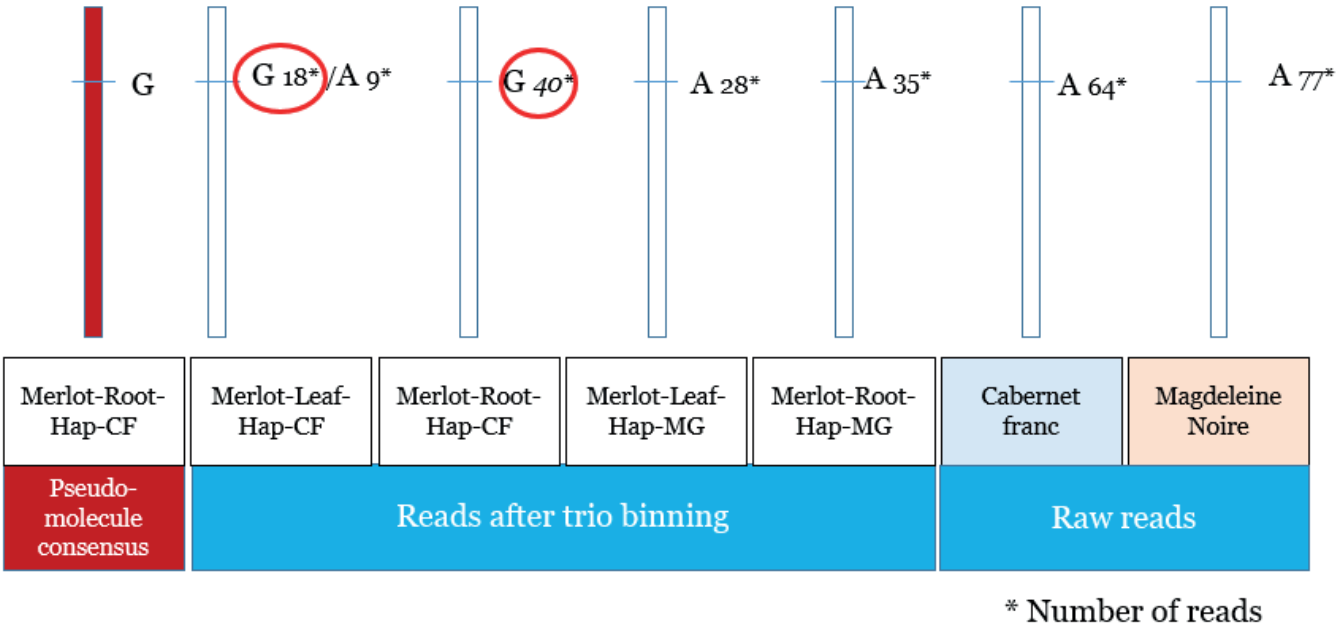

Fig. 2: An example of read distribution when a L2 chimera was detected.

Table 1: Chromosome length, contig numbers and gene numbers for each pseudomolecule

	Merlot_leaf_CF			Merlot_leaf_MG			Merlot_root_CF			Merlot_root_MG		
	Length	Contigs	Genes	Length	Contigs	Genes	Length	Contigs	Genes	Length	Contigs	Genes
chr01	23598264	2	2091	25285196	3	2308	23589260	2	2090	25812159	3	2229
chr02	21618960	2	1711	20834059	2	1771	21612554	2	1710	20815648	2	1773
chr03	22910969	2	1896	23738413	3	1853	22695420	4	1900	22863606	2	1856
chr04	28028847	1	2171	28067391	1	2199	27848422	1	2172	28047555	1	2195
chr05	28252839	3	2354	27194848	4	2190	28133326	2	2359	27245740	5	2197
chr06	21902219	1	1997	25066429	2	1901	21887941	2	1999	25064422	2	1900
chr07	31335031	1	2892	30285517	2	2877	31316563	1	2889	30272463	2	2875
chr08	25084147	2	2260	25133383	2	2159	25068184	1	2264	25106757	2	2163
chr09	23385614	3	1774	25283510	3	1789	23388363	3	1772	25079599	3	1785
chr10	26473290	1	2156	25356578	3	2159	26466742	1	2159	25373432	3	2160
chr11	20347376	2	1571	20416243	1	1563	19944127	2	1572	20402067	1	1569
chr12	27064986	2	2447	24119349	4	2347	24025996	1	2400	24096300	3	2343
chr13	29400470	2	2277	29181387	2	2244	29412682	3	2277	29182803	3	2245
chr14	30137549	1	2577	31130771	2	2564	30138415	1	2570	31044170	1	2545
chr15	23518695	2	1552	23434760	2	1832	23494571	2	1545	23422104	2	1828
chr16	22496135	1	1813	22861382	3	1840	22480932	3	1815	22716031	2	1844
chr17	20156470	2	1552	20678794	5	1628	20368439	3	1554	20784459	4	1628
chr18	36729617	2	3197	37876454	1	3161	36718487	2	3192	37875940	1	3164
chr19	27613142	1	1978	25782670	2	1992	27618512	2	1979	25770078	2	1994
Total length	490054620	33	40266	491727134	47	40377	486208936	38	40218	490975333	44	40293
Total + UKN	499510259			499510259			500119474			503747540		
chrUn	9455639			7783125			13910538			12772207		
% Total length	1.90%			1.60%			2.80%			2.50%		

Table 2: Number of chimeras found on each Merlot haplotypes and their distribution according to coding regions on the genome

	Merlot Haplotype CF			Merlot Haplotype MG		
	L1	L2	L1 + L2	L1	L2	L1 + L2
SNP, Non-repeated sequences Merlot specific periclinal chimeras	37	14	**51**	36	17	**53**
Number of chimeras in non coding gene body	15	4	**19**	11	5	**16**
Number of chimeras in coding region	6	3	**9**	6	1	**7**

to intra-varietal genetic diversity. Finally, if they are stable enough, they could be used as a clonal lineage signature and allow clonal identification.

Additional data can be obtained from the following preprint in BMC Genomics: Sichel, V., Sarah, G., Girollet, V., Laucou, V., Roux, C., Bert, P., Le Cunff, L., This, P., Lacombe, T., 2022: Chimeras in Merlot grapevine revealed by phased assembly. https://doi.org/10.21203/rs.3.rs-2026816/v1

Conflicts of interest

The authors declare that they do not have any conflicts of interest.

References

Bauhin, C., 1598: Illustrated Exposition of Plants (the Pinax theatri botanici). sumptibus [et] typis Ludovici Regis.

Blakeslee, A.F., Avery, A.G., Bergner, A.D., Satina, S.A., Sinnott, E.W., 1939: Induction of periclinal chimeras in Datura stramonium by colchicine treatment. Science 89 (2314), 402.

Boursiquot, J.M., Lacombe, T., Laucou, V., Julliard, S., Perrin, F.X., Lanier, N., Legrand, D., Meredith, C., This, P., 2009: Parentage of Merlot and related winegrape cultivars of southwestern France: discovery of the missing link. Australian Journal of Grape and Wine Research 15 (2), 144-155, DOI: 10.1111/j.1755-0238.2008.00041.x.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., Li, H., 2021: Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature Methods 18 (2), 170-175, DOI: 10.1038/s41592-020-01056-5.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., Depristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011: The variant call format and VCFtools. Bioinformatics 27 (15), 2156-2158, DOI: 10.1093/bioinformatics/btr330.

Nati, P., 1644: Petri Nati... Florentina phytologica obseruatio de malo limonaia citrata-aurantia Florentiae vulgo la bizzarria. Typis Hippolyti de Naue.

Pratt, C., Einset, J., Zahur, M., 1959: Radiation Damage in Apple Shoot Apices. American Journal of Botany 46 (7), 537-544, DOI: 10.2307/2439626.

Rustenholz, C., Velt, A., Frommer, B., 2021: PN40024.v4 New Assembly. URL: https://integrape.eu/resources/genes-genomes/genome-accessions/.

Ryan Poplin, P.C.C., David Alexander, Scott Schwartz, Thomas Colthurst, Alexander Ku, Dan Newburger, Jojo Dijamco, Nam Nguyen, Pegah T. Afshar, Sam S. Gross, Lizzie Dorfman, Cory Y. McLean, and Mark A. DePristo, 2018: DeepVariant: A universal SNP and small-indel variant caller using deep neural networks. Nature biotechnology 36, 983-987, DOI: 10.1038/nbt.4235.

Satina, S., 1945: Periclinal chimeras in Datura in relation to the development and structure of the ovule. American Journal of Botany, 72-81.

Shumate, A., and Steven L. Salzberg, 2021: Liftoff: accurate mapping of gene annotations. Bioinformatics 37 (12), 1639-1643, DOI: 10.1093/bioinformatics/btaa1016.

Smit, A.F.A., Hubley, R., Green, P., 2013-2015: RepeatMasker. URL: http://repeatmasker.org.

Szymkowiak, E.J., Sussex, I.M., 1996: What chimeras can tell us about plant development. Annual Review of Plant Biology 47 (1), 351-376.

Thompson, M.M., Olmo, H., 1963: Cytohistological studies of cytochimeric and tetraploid grapes. American Journal of Botany 50 (9), 901-906.

Torregrosa, L.J.M., Rienth, M., Romieu, C., Pellegrino, A., 2019: The microvine, a model for studies in grapevine physiology and genetics. OENO One 53 (3), DOI: 10.20870/oeno-one.2019.53.3.2409.

Vezzulli, S., Leonardelli, L., Malossini, U., Stefanini, M., Velasco, R., Moser, C., 2012: Pinot blanc and Pinot gris arose as independent somatic mutations of Pinot noir. Journal of Experimental Botany 63 (18), 6359-6369, DOI: 10.1093/jxb/ers290.