

Genomics technologies to study structural variations in the grapevine genome

Maria Francesca Cardone¹, Carlo Bergamini¹, Pietro D'Addabbo², Can Alkan³, Claudia Rita Catacchio², Fabio Anaclerio², Giorgia Chiatante^{1,2}, Annamaria Marra², Giuliana Giannuzzi², Rocco Perniola¹, Mario Ventura², and Donato Antonacci¹

¹ Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA), Center for viticulture and enology, 70010 Turi (BA), Italy

² Università degli Studi di Bari "Aldo Moro", Dipartimento di Biologia, 70126 Bari, Italy

³ Bilkent University, Department of Computer Engineering, TR-06800, Bilkent, Ankara, Turkey

Abstract. Grapevine is one of the most important crop plants in the world. Recently there was great expansion of genomics resources about grapevine genome, thus providing increasing efforts for molecular breeding. Current cultivars display a great level of inter-specific differentiation that needs to be investigated to reach a comprehensive understanding of the genetic basis of phenotypic differences, and to find responsible genes selected by cross breeding programs. While there have been significant advances in resolving the pattern and nature of single nucleotide polymorphisms (SNPs) on plant genomes, few data are available on copy number variation (CNV). Furthermore association between structural variations and phenotypes has been described in only a few cases. We combined high throughput biotechnologies and bioinformatics tools, to reveal the first inter-varietal atlas of structural variation (SV) for the grapevine genome. We sequenced and compared four table grape cultivars with the Pinot noir inbred line PN40024 genome as the reference. We detected roughly 8% of the grapevine genome affected by genomic variations. Taken into account phenotypic differences existing among the studied varieties we performed comparison of SVs among them and the reference and next we performed an in-depth analysis of gene content of polymorphic regions. This allowed us to identify genes showing differences in copy number as putative functional candidates for important traits in grapevine cultivation.

1. Introduction

It is nowadays demonstrated that the genomes of two individuals of the same species can exhibit substantial genetic variation both in term of sequence variation and structure alteration. Discovery and characterization of all forms of these genetic variations are crucial to reach a comprehensive understanding of the genetic basis of phenotypic differences.

Grapes and their derivatives have a large and expanding worldwide market, not only as wine, fresh fruit, or raisins, but also for its recent application in human health and cosmetics. The grape has great potential in plant biology to become a model organism for fruit lineages as it can be transformed and micro-propagated via somatic embryogenesis [1]. Two genotypes derived from Pinot noir cultivars (wine grape) have been sequenced and assembled as reference genomes [1,2] and later the genome of the Thompson seedless cultivar (table grape) was sequenced and assembled [3]. Most modern grape varieties are the final effect of human selection and vegetative reproduction, aiming to isolate specific traits in terms of pathogenic resistance or crop production. In this way, current cultivars display a great level of inter-specific differentiation that needs to be investigated to find responsible genes selected by cross-breeding programs. In this regard, there is now great interest

in the genomic variation in grape, such as single nucleotide variants (SNVs, e.g. SNPs), small INDELS, and structural variations (SVs), which include large copy number variation (CNV). While SNVs can be simply identified through genome comparison and generally show easy-to-detect effects on gene function, SVs—such as large duplications and deletions—need more effort to be correctly characterized. Besides the importance of CNVs, our understanding of the most prevalent contributors to CNVs in plants is still far from being well explored. Recently, many methods to discover genomic variations have been developed and primarily applied to human, primates, and other mammalian genomes [4–6].

In the present paper we described the first map of structural variation (SV) in the grapevine genome. This paper summarizes some of the results reported in Cardone et al. 2016 [7].

We combined high throughput sequencing (HTS) with array comparative genomic hybridization (CGH), fluorescent in situ hybridization (FISH), and quantitative PCR to create the first comprehensive map of genomic variations in four table grape genomes. We sequenced and compared the four table grape cultivars—Autumn royal (AR), Italia (It), Red globe (RG) and Thompson seedless (TS)—with the Pinot noir genome as the reference (wine grape inbred line PN40024). We found an average of 8% of the grapevine genome is affected by genomic variations,

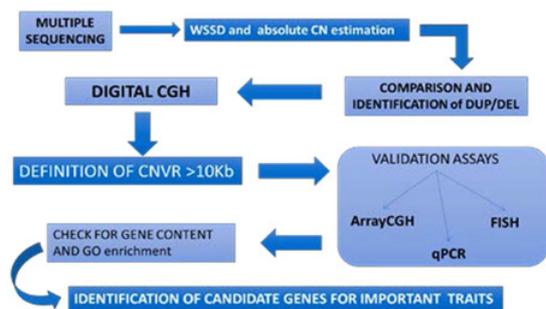


Figure 1. Schematic representation of the workflow used in the analysis.

and we were able to identify inter- varietal-specific CNVs. Next we performed an in-depth analysis of gene content of polymorphic regions. This allowed us to identify genes and/or gene families as functional candidates for important traits in grapevine cultivation that can be used as genetic tools for breed selection programs. Overall data could represent a landmark for future comparative studies and it could be considered as a step towards the definition of a grape “pan-genome”.

2. Material and methods

Taken into account data on human, mammals, and grapevines [8–10] we combined modern *high-throughput* technologies to develop a new approach for plant genome studies to describe the genomic structure of multiple genomes at the same time. We paired-end sequenced four table grape cultivars: Autumn royal (AR), Italia (It), Red globe (RG), and Thompson seedless (TS) with a coverage ranging from 13X to 19X and alignment of the obtained reads against the PN40024 Pinot noir reference genome. Further, we defined a specific workflow to identify duplication/deletions and CNVs in four table grape varieties in comparison to the reference genome of Pinot noir (inbreed line PN40024) as described in the Fig. 1. For additional detailed methods see [7, 10].

In summary:

DUP/DEL identification. After pair end sequencing, we defined the segmental duplication (SD) content in the four genomes using a version of the WSSD approach modified for HTS data [8]. In particular for the first time on a plant genome we applied a method that allowed the estimation of the absolute CN counts of each region. We used mrCaNaVaR (micro-read Copy Number Variant Regions), a copy number caller able to predict absolute copy number, and analyzed the whole-Genome NGS data. This algorithm leads to the detection of duplicated and deleted genomic regions, highlighted by a local excess of depth of coverage or a local reduction of depth of coverage, respectively.

Digital CGH. WSSD data were compared between each of the sequenced varieties vs. PN40024 and among them with all the different possible combination to identify common and unique SDs or DELs. CN of each window and regions were calculated and “digital CGH” as reported by Sudmant et al. [4] with some modification, was performed to found CNV Regions (CNVRs) >10kb. In particular, the genome of each variety was first masked to filter out repeats, and the unmasked genome was then

divided into regions of 1 Kb and the copy number of these regions were compared by using the following formula:

$$L2R = \text{Log}_2(\text{CN var.genome}/\text{CN ref.genome}) \quad (1)$$

Validation. As there are no previous studies on genome-wide discovery of CNVRs in grapevine, and only a few other plant genomes were analyzed for CNVRs using similar approaches, different tests were performed in order to choose the best criteria to find bigger and significant CNVs. Moreover, predicted CNVRs were validated by array CGH and FISH assay. In addition quantitative PCR was performed to further validate the CN estimation for a selection of regions.

Functional analysis. We classified the CNVRs based on common or unique phenotypes of the four varieties. Then gene content and relative annotation were reported for each class and for each identified CNVR. We checked for functional category enrichment. Based on the reported gene ontology and functional annotation, we classified CNVRs based on their gene content and searched for possible candidate genes. CNVRs containing functional candidates were further investigated.

3. Results and discussion

3.1. Genomic map of plastic regions and CNVs in the grapevine genome

Previously the same research group has studied the plasticity in the grapevine genome by using bioinformatics algorithms based on the higher depth of coverage of WGS sequence reads, aligned to the reference genome sequence, and revealed that 85 Mb of the grapevine reference genome were segmentally duplicated (17%) [10]. Further, by using InterProScan tools, we investigated the gene content of the 100 most duplicated regions and found enrichment in genes coding for Cytochrome P450, a key enzyme involved in the biosynthesis of several compounds such as hormones, defensive compounds and fatty acids; genes coding for enzymes involved in secondary metabolisms, and genes involved in immune response, xenobiotic recognition, reproduction and nuclear functions. These data have already demonstrated the important role of segmental duplication in adaptive evolution. Moreover segmental duplications represent one of the greatest source of structural variation.

For this reason we decided to further study these regions and to improve the previously used approach to investigate not only segmental duplication, but also deletions and most important structural variation in the grapevine genome.

We then combined modern *high-throughput* technologies to develop a new approach for plant genome studies, and by comparing the genome of four table grape varieties (AR, It, RG, TS) with respect to the reference genome of PN40024, we defined the first map of structural variation in the grapevine genome.

By our genome-wide analysis, conducted both on the four table grapes and the PN40024 reference genome, we revealed that deletions and highly identical SDs characterize about 9% and 26% of grape genome, respectively. Notably, SDs and deletions that are common to the four table grapes characterized only 1.72% and 0.55% of the genome, respectively.

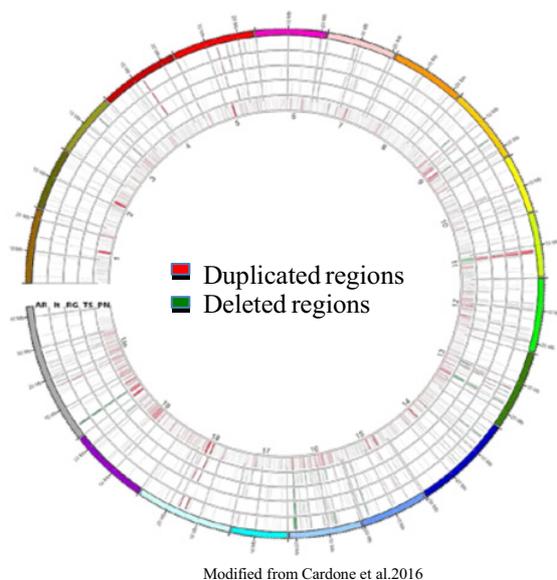


Figure 2. Circular representation of dCGH data in four varieties of grapes. Deleted (green) and duplicated (red) regions, detected by dCGH with respect to the PN40024 genome, were graphically highlighted in four circular representations of the genomes of the analyzed grapes varieties. The external colored circle represents the 19 *Vitis vinifera* L. chromosomes and the “unknown” chromosome (that collects regions sequenced but unassigned to a specific chromosome). Chromosome name is reported and vertical gray lines delimit the start and end of each chromosome. The internal circles tag CNVs found in each region, reported in order, in Autumn royal (AR), Italia (It), Red globe (RG), and Thompson seedless (TS). The inner circle represents the WSSD map in the PN40024 reference genome.

Focusing on gene content of regions common to the table grapes genomes, we found that genes involved in stress responses, such as *NBS-LRR* genes and those related to qualitative aspects of berries (e.g., genes involved in flavonoid or terpenoid metabolism), underwent additional duplication events. These genes have been already described as the primary target of duplication events in the grapevine genome [10].

Nevertheless, the finding of SDs specific to table grapes with respect to the reference genome deserves future investigation to identify genes specifically involved in table grape quality characteristics.

Next, to detect CNVs differentiating the varieties, we applied an in silico digital CGH approach on the whole genome similar to an algorithm described to characterize CNVs within human genomes [4].

We found a total of 746 CNVRs (>10 kbp) across the four varieties: 310 in It, 318 in RG, 355 in TS, and 350 in AR (Fig. 2). This corresponds to a percentage of the genome, ranging from 3.35% in It to 4.05% in TS, which are equally distributed between gains and losses of paralogous copies.

In each of the four varieties, about 35% of these regions were large CNVs greater than 50 kbp, and 10% were greater than 100 kbp. 335 CNVs out of 746 are uniquely identified in one variety, while 64 were found in all the four table grape genomes with respect to the reference. Notably, 46% of the CNVRs were mapped in plastic regions (both duplicated or deleted) and in particular about 41% of CNVs mapped in regions duplicated in the

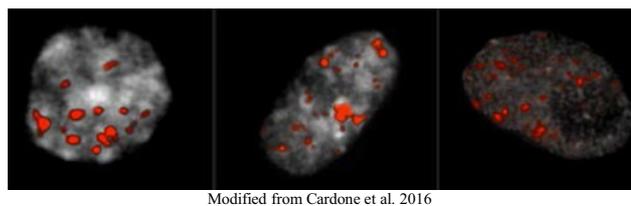


Figure 3. Exemplifying FISH patterns of a BAC clone mapping in an hyper duplicated and highly variable region, on interphase nuclei obtained from Thompson seedless, Red globe, and Italia leaf buds (from right).

Table 1. Comparison between aCGH and dCGH. Data are reported as percentage of *Vitis vinifera* L. genome.

	AR	IT	RG	TS
CNVs detected by aCGH	2.03	1.93	1.71	2.08
CNVs detected by dCGH	4.03	3.36	3.69	4.06
CNVs detected by both approaches	1.21	0.73	1.13	1.24
Not polymorphic regions detected by both approaches	95.15	95.44	95.73	95.1
Total concordance rate (previous two rows)	96.36	96.17	96.86	96.34

reference genome, while 5% matched with regions deleted in the PN40024 genome. Overall comparison showed that CNV affected about 8% of grapevine genome. These data suggest that the entire grape genome is highly dynamic and subject to structural alterations and this genomic variability could reflect the great phenotypic diversity existing in the *Vitis* genus. The high number of CNVs found in such a small genome supports the importance of structural variations in shaping the grapevine genomes.

Due to the novelty of the used approach we performed three different validation assays.

FISH assay: 43 BAC clones were tested on interphase nuclei derived from the sequenced varieties. The validation rate was variable among the varieties, ranging from 58% in TS, 63% in RG, to 68% in AR and It. The analysis of the FISH patterns revealed the existence of regions that are hyper duplicated and highly variable among the studied *Vitis vinifera* cultivars (Fig. 3).

ArrayCGH: Array CGH revealed CNVs in about 2% of grape genome for each variety. In particular, as both array CGH and digital CGH are genome-wide approaches, we checked the correspondence of the calls performed by array and digital CGH on the whole genome, and not only on CNV putative regions. We considered a prediction consistent if a region was called as a CNV or as not polymorphic at the same time by both methods. As described in Table 1, we found a high level of correspondence (>96%) between array CGH and digital CGH.

qPCR assay: Primers pairs were designed to confirm 21 genic regions predicted to be polymorphic among the four sequenced varieties. In particular, we selected three regions predicted as constantly diploid and 18 predicted

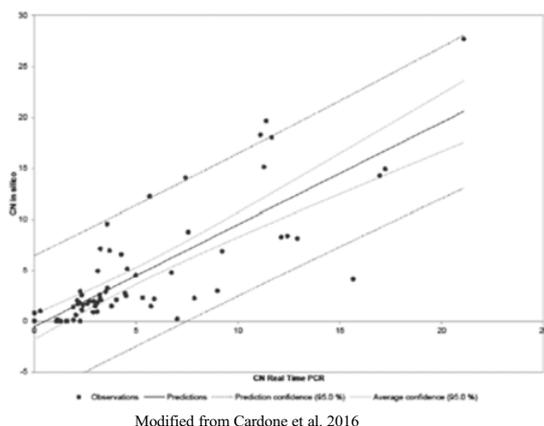


Figure 4. Regression curve obtained by the comparison between the CN calculated by in silico approach and CN derived by qPCR.

to be highly polymorphic among the varieties. CN was estimated using the relative standard curve method, comparing to an endogenous reference gene arbitrarily taken as constantly diploid, the fructose-6-phosphate-2-kinase. Among the 21 tested genes, 17 were validated by qPCR assays. As a further corroboration of the reliability of the absolute CN in silico predictions, we calculated the linear regression among all the CN in silico with respect to the CNqPCR (Fig. 4). The function describing the regression of the data was found as follows:

$$\text{CN in silico} = 1.001 * \text{CNqPCR} - 0.540. \quad (2)$$

This latter result revealed the consistency of the absolute CN in silico prediction.

3.2. Polymorphic genes

In addition to the definition of the CNV map, for each region we assessed the CN and gene content. In order to find polymorphic genes, we compared the CNVRs with *Vitis vinifera* L. gene annotation and we searched for specific functional category enrichment among all polymorphic regions.

Half of the CNVRs were found overlapping SDs. This confirms that SDs are hotspots for CNV formation [1,8,11] and that the non-allelic homologous recombination could be one of the primary CNV rising mechanisms in plants and likewise in human. We also found many transposable elements among polymorphic regions, especially among the deleted regions. This result is in agreement with that found in *Arabidopsis thaliana* [13] and, thus, supports the important role of transposons movement in mediating deletions and SV onset [14,15].

As a general overview, in agreement with data from works exploring genomes of other plant species [16,17], most of the polymorphic genes we detected belong to some well-known gene super-families such as the MYB transcription factors (involved in the synthesis of molecules, which confer quality to grapes and wines), the TPS genes [18] (involved in flavor determination) or NBS-LRR [19] (stress response gene families). CNVs in such genes could explain the different adaptations to respond to external environmental stresses of one variety with respect

to another. As a future perspective, specific polymorphisms found in these genes could be useful in molecular breeding for their stress resistance.

Interestingly, considering sub-categorization of polymorphisms, we highlighted the importance of CNVs in grapevine and identified candidate genes for some of the most complex and desired genetically selected traits in breeding programs. For instance, considering CNVs common to the four table grapes, we found a CNV in a gene that belongs to the expansin family. In particular, we found that a gene of the EXPA4 family underwent additional amplification in table grapes with respect to the reference genome. Recently, genome-wide analyses and expression data of these genes revealed they are finely modulated during fruit growth and maturation and, thus, have an important role in processes critical to determining berry quality [20]. The economic importance of grapevine is greatly influenced by the quality of its berry. This is especially true for table grape as berries are the final product.

Seedlessness and berry size are among the most studied traits in table grapes as principal targets in molecular breeding programs. An undesired negative correlation exists between seedlessness and berry size since seed tissues supply important hormones for fruit development [21]. This aspect is probably under a strict hormonal control, but it is not (at least not only) due to a deficiency in plant growth-regulator levels; on the contrary, it is more likely associated with the minor quantitative trait loci involved in the seedlessness and/or a combination of different quantitative trait loci. Comparing polymorphic regions between seedless (AR and TS) vs. seeded (It and RG) and big berry vs. small berry (TS) varieties, we were able to identify polymorphisms in genes involved in hormone responses and metabolism in berry growth as putative candidate genes. For example the gene coding for the Auxin Response Factor 5 (ARF5) showed a higher CN in the TS genome and was recently mapped to a quantitative trait locus associated to berry weight and traits [22]. Likewise we found polymorphisms specific to the seedless varieties with respect to the seeded ones in genes involved in transport overview pathways, such as PIP2B, which codes a member of the aquaporin gene family.

Similarly among the studied varieties the Italia is the only one presenting aromatic characteristics. Thus we focus our attention on polymorphic genes specific in the Italia genome and we found that the gene for the germacrene D synthase belonging to the TPS superfamily, was amplified in all the analyzed varieties but showed a higher CN in Italia. CNVs in TPS could be related to differences in the metabolic pathways of this compound and contribute to differences in the aromatic flavor of one variety with respect to another one. In this context, the polymorphisms found in the germacrene D synthase and other TPS genes represent good candidate genes and deserve further investigation.

4. Conclusion

In the present paper we described the first inter-varietal atlas of CNVs in *Vitis vinifera* L.

SVs in plants have been previously considered to be part of the so-called “dispensable genome” and not

necessary for survival [14]. Nevertheless, recent studies on their importance revealed that the distinction between core and dispensable genome is not immutable and SVs could be considered as “conditionally dispensable” [15]. Our data support this important role of SVs in the grapevine genome.

Taken together, our data assess that plastic regions represent more than 26% of the grapevine genome and 8% is variant among different varieties.

We also performed an in deep analysis of the gene content of CNVs in order to find putative candidate genes for important phenotypic traits. We were able to detect varietal CNVs in genes involved in aromatic compound biosynthesis and metabolism. Likewise, we found notable genomic variation differences for genes playing critical roles in stress response to both biotic and abiotic stresses.

Finally, still a lot is missing! Almost 40% of the regions detected as CNVs showed no functional gene annotation or predicted to produce unknown/unclear proteins. These regions deserve further studies to understand genes function and improve the gene annotation in the grapevine genome.

References

- [1] R. Velasco, A. Zharkikh, M. Troglio, D. A. Cartwright, A. Cestaro, D. Pruss, M. Pindo, L. M. Fitzgerald, S. Vezzulli, J. Reid, G. Malacarne, D. Iliev, G. Coppola, B. Wardell, D. Micheletti, T. Macalma, M. Facci, J. T. Mitchell, M. Perazzolli, G. Eldredge, P. Gatto, R. Oyzerski, M. Moretto, N. Gutin, M. Stefanini, Y. Chen, C. Segala, C. Davenport, L. Dematt, A. Mraz, J. Battilana, K. Stormo, F. Costa, Q. Tao, A. Si-Ammour, T. Harkins, A. Lackey, C. Perbost, B. Taillon, A. Stella, V. Solovyev, J. A. Fawcett, L. Sterck, K. Vandepoele, S. M. Grando, S. Toppo, C. Moser, J. Lanchbury, R. Bogden, M. Skolnick, V. Sgaramella, S. K. Bhatnagar, P. Fontana, A. Gutin, Y. Van de Peer, F. Salamini, and R. Viola. *PLoS One* **2**, e1326 (2007)
- [2] O. Jaillon J. M. Aury, B. Noel, A. Policriti, C. Clepet, A. Casagrande, N. Choisne, S. Aubourg, N. Vitulo, C. Jubin, A. Vezzi, F. Legeai, P. Huguency, C. Dasilva, D. Horner, E. Mica, D. Jublot, J. Poulain, C. Bruyère, A. Billault, B. Segurens, M. Gouyvenoux, E. Ugarte, F. Cattonaro, V. Anthouard, V. Vico, C. Del Fabbro, M. Alaux, G. Di Gaspero, V. Dumas, N. Felice, S. Paillard, I. Juman, M. Moroldo, S. Scalabrin, A. Canaguier, I. Le Clainche, G. Malacrida, E. Durand, G. Pesole, V. Laucou, P. Chatelet, D. Merdinoglu, M. Delledonne, M. Pezzotti, A. Lecharny, C. Scarpelli, F. Artiguenave, M. E. Pè, G. Valle, M. Morgante, M. Caboche, A. F. Adam-Blondon, J. Weissenbach, F. Quérier, P. Wincker, and French-Italian Public Consortium for Grapevine Genome Characterization. *Nature* **449**, 463–467 (2007)
- [3] A. Di Genova, A. M. Almeida, C. Muñoz-Espinoza, P. Vizoso, D. Travisany, C. Moraga, M. Pinto, P. Hinrichsen, A. Orellana, and A. Maass. *BMC Plant Biol* **14**, 7 (2014)
- [4] P. H. Sudmant, J. O. Kitzman, F. Antonacci, C. Alkan, M. Malig, A. Tsalenko, N. Sampas, L. Bruhn, J. Shendure, E. E. Eichler, and 1000 Genomes Project. *Science* **330**, 641–646 (2010)
- [5] C. Alkan, B. P. Coe, E. E. Eichler. *Nat Rev Genet* **12**, 363–376 (2011)
- [6] R. E. Mills, K. Walter, C. Stewart, R. E. Handsaker, K. Chen, C. Alkan, A. Abyzov, S. C. Yoon, K. Ye, R. K. Cheatham, A. Chinwalla, D. F. Conrad, Y. Fu, F. Grubert, I. Hajirasouliha, F. Hormozdiari, L. M. Iakoucheva, Z. Iqbal, S. Kang, J. M. Kidd, M. K. Konkel, J. Korn, E. Khurana, D. Kural, H. Y. Lam, J. Leng, R. Li, Y. Li, C. Y. Lin, R. Luo, X. J. Mu, J. Nemes, H. E. Peckham, T. Rausch, A. Scally, X. Shi, M. P. Stromberg, A. M. Stütz, A. E. Urban, J. A. Walker, J. Wu, Y. Zhang, Z. D. Zhang, M. A. Batzer, L. Ding, G. T. Marth, G. McVean, J. Sebat, M. Snyder, J. Wang, E. E. Eichler, M. B. Gerstein, M. E. Hurles, C. Lee, S. A. McCarroll, J. O. Korb, and 1000 Genomes Project. *Nature* **470**, 59–65 (2011)
- [7] M. F. Cardone, P. D’Addabbo, C. Alkan, C. Bergamini, C. R. Catacchio, F. Anaclerio, G. Chiatante, A. Marra, G. Giannuzzi, R. Perniola, M. Ventura, and D. Antonacci. *Plant J*, in press (2016)
- [8] C. Alkan, M. Kidd, T. Marques-Bonet, G. Aksay, F. Antonacci, F. Hormozdiari, J. O. Kitzman, C. Baker, M. Malig, O. Mutlu, S. C. Sahinalp, R. A. Gibbs, and E. E. Eichler. *Nat Genet* **41**, 1061–1067 (2009)
- [9] D. M. Bickhart, Y. Hou, S. G. Schroeder, C. Alkan, M. F. Cardone, L. K. Matukumalli, J. Song, R. D. Schnabe, M. Ventura, J. F. Taylor, J. F. Garcia, C. P. Van Tasse, T. S. Sonstegard, E. E. Eichler, and G. E. Liu. *Genome Res* **22**, 778–790 (2012)
- [10] G. Giannuzzi P. D’Addabbo, M. Gasparro, M. Martinelli, F. N. Carelli, D. Antonacci, and M. Ventura. *BMC Gen.* **12**, 436 (2011)
- [11] A. J. Sharp D. P. Locke, S. D. McGrath, Z. Cheng, J. A. Bailey, R. U. Vallente, L. M. Pertz, R. A. Clark, S. Schwartz, R. Segraves, V. V. Oseroff, D. G. Albertson, D. Pinkel, and E. E. Eichler. *Am J Hum Genet* **77**, 78–88 (2005)
- [12] T. Marques-Bonet J. M. Kidd, M. Ventura, T. A. Graves, Z. Cheng, L.D. W. Hillier, Z. Jiang, C. Baker, R. Malfavon-Borja, L. A. Fulton, C. Alkan, G. Aksay, S. Girirajan, P. Siswara, L. Chen, M. F. Cardone, A. Navarro, E. R. Mardis, R. K. Wilson, and E. E. Eichler. *Nature* **457**, 877–881 (2009)
- [13] S. DeBolt. *Genome Biol Evol* **2**, 441–453 (2010)
- [14] M. Morgante, E. De Paoli, S. Radovic. *Curr Opin Plant Biol* **10**, 149–155 (2007)
- [15] F. Marroni, S. Pinosio, M. Morgante. *Curr Opin Plant Biol* **18**, 31–36 (2014)
- [16] R. A. Swanson-Wagner, S. R. Eichten, S. Kumari, P. Tiffin, J. C. Stein, D. Ware, and N. M. Springer. *Genome Res* **20**, 1689–1699 (2010)

- [17] M. Muñoz-Amatriaín *et al.*, Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. *Genome Biol* **14**, R58 (2013)
- [18] D. M. Martin, S. Aubourg, M. B. Schouwey, L. Daviet, M. Schalk, O. Toub, S. T. Lund, and J. Bohlmann. *BMC Plant Biol* **10**, 226 (2010)
- [19] G. Malacarne, M. Perazzolli, A. Cestaro, L. Sterck, P. Fontana, Y. Van de Peer, R. Viola, R. Velasco, and F. Salamini. *PLoS One* **7**, e29762 (2012)
- [20] S. Dal Santo, A. Vannozzi, G. B. Tornielli, M. Fasoli, L. Venturini, M. Pezzotti, and S. Zenoni. *PLoS One* **8**, e62206 (2013)
- [21] G. Fanizza, F. Lamaj, L. Costantini, R. Chaabane, M. S. Grando. *Theor Appl Genet* **111**, 658–664 (2005)
- [22] A. Doligez, Y. Bertrand, M. Farnos, M. Grolier, C. Romieu, F. Esnault, S. Dias, G. Berger, P. François, T. Pons, P. Ortigosa, C. Roux, C. Houel, V. Laucou, R. Bacilieri, J. P. Péros, and P. This. *BMC Plant Biol* **13**, 217 (2013)