

# Bioinformatic Tools for Polyploid Crops

Fabian Grandke<sup>1,2</sup>, Soumya Ranganathan<sup>2</sup>, Andrzej Czech<sup>1</sup>, Jorn R. de Haan<sup>1</sup> and Dirk Metzler<sup>2</sup>

1. Bioinformatics Department, Genetwister Technologies B. V., Wageningen 6700, Netherlands

2. Department of Biology, University of Munich, Planegg-Martinsried 82512, Germany

Received: July 7, 2014 / Published: August 20, 2014.

**Abstract:** Polyploidy is common among agriculturally important crops. Popular genetic methods and their implementations cannot always be applied to polyploid genetic data. We give an overview about available tools and their limitations in terms of levels of ploidy, auto- and allo-ploidy. The main classes of tools are genotype calling, linkage mapping and haplotyping. The usability of the tools is discussed with a focus on their applicability to data sets produced by state of the art technologies. We show that many challenges remain until the toolset for polyploidy provides similar functionalities as those which are already available for diploids. Some tools have been developed over a decade ago and are now outdated. In addition, we discuss necessary steps to overcome this shortage in the future.

**Key words:** Polyploid, molecular breeding genotyping, haplotyping, linkage mapping, bioinformatics, quantitative genetics, phasing.

## 1. Introduction

Polyploidy describes the state of having more than two sets of chromosomes in a cell. Different levels of ploidy are possible, but even-numbered ploidies and especially tetraploidy (four sets of chromosomes) are the most common ones [1]. In general, polyploid species are rare, but frequent in flowering plants (angiosperms). Particularly plant species that are important to horticulture and agriculture are polyploids (e.g., potato, sugarcane and brassica). In some species multiple ploidy levels exist [2]. Nevertheless, polyploidy is not unique to plants. Several animals are known to be polyploid. The ploidy levels range from three in various reptiles to twelve in the Uganda clawed frog (*Xenopus ruwenzoriensis*) [3, 4]. In plants, the additional chromosomes cause changes in a range of traits. Those can be either advantageous or disadvantageous for the plant's fitness [1]. One of them is the gigas

effect, which describes the increase in the cell size of polyploid plants in comparison to diploid individuals of the same species [5]. It leads to increase plant organ sizes (e.g., larger fruits), which is highly appreciated [6]. Polyploids can have a decreased growth rate, compared to diploids [5]. This leads to later and longer flower seasons, which are of interest for plant breeders. Other affected traits include reduced fertility, different chemical composition and many more.

On the long term polyploidy is disadvantageous for an individual's fitness and the genomes tend to diploidize in nature [1]. Besides the natural spontaneous origin, polyploidy can be induced artificially [7, 8]. Polyploids can be classified as allopolyploid, autopolyploid or segmental allopolyploids. Allopolyploidy describes that the increased chromosome sets originate from multiple species, while autopolyploids consist of multiple copies of the same chromosome set [5]. In general, polyploids are treated similar to diploids. The same technologies and tools are used for genome sequencing, assembly, read mapping and single nucleotide polymorphism (SNP) detection. This strategy fails, as soon as a genotype analysis is

---

**Corresponding author:** Fabian Grandke, M.Sc., research fields: crop bioinformatics, polyploid crops and genotype/phenotype associations. E-mail: f.grandke@genetwister.nl.

involved, because the genotypes of polyploids have larger dimensions than diploids.

Genetic markers are indicators for genetic variation. Their sizes and complexity range from single nucleotides to long stretches of DNA. The comparison of the markers in multiple individuals or populations can provide information about the relation between them. In addition, the markers can be used to characterize the genotype of an individual. Nowadays, SNPs are the state-of-the-art genetic markers in genetics [9]. They provide a high density over the genome and are used in plant and animal breeding, as well as in medicine. Theoretically, a SNP can have up to four different variants (A, C, G and T), but most SNPs are biallelic, i.e., only two nucleotides are common at their position [10]. A SNP marker in a diploid (allele A and B) can have three different variations: AA, AB and BB. Polyploids have increased numbers of possible genotypes, relative to their ploidy level. For instance, a tetraploid has five: AAAA, AAAB, AABB, ABBB and BBBB. Thus, tools and formats initially developed for diploids need to be extended to accept the additional genotype variants.

Quantitative genetics is the underlying theory of modern breeding [11]. It relies on genetic markers and most of its implemented tools need to be adapted for polyploids. The basic concept is that phenotypes depend on genotypes and environmental factors. Usually a genotype is represented by the combinations of its genetic markers. Quantitative genetics uses statistical methods to associate genotypes with phenotypes in order to obtain information about genes. In absence of genetic maps, the recombination between markers is used to calculate a linkage map [12]. This can be a tool to mark regions in a genome linked to quantitative traits (quantitative trait loci, QTL).

In this manuscript, we give an overview of tools available for polyploid genetic data analysis. We discuss what methods are lacking and still need to be

developed to allow for quantitative genetic breeding in polyploids. This is important to perform advanced genomic breeding to improve economically important traits like disease resistances or yield. Lately, the polyploid genetic toolbox has increased strongly. Hence, even for complex genomes, state of the art breeding techniques can be applied now. With this overview, we want to inform breeders about the latest developments in the field and encourage developers to account for polyploids.

## 2. The Polyploid Toolbox

We consider tools for three different problems to be solved (also) for polyploid data:

- (1) Genotype callers
- (2) Linkage mappers
- (3) Haplotype phasers

All of them have been developed for diploids and are established in research. To be applicable to polyploids, they require extensions, modifications or new approaches [12].

### 2.1 Genotype Calling

To apply the marker information, it is necessary to know the alleles of each marker for each sample. Nevertheless, missing values are possible. This can be done by the use of genotyping microarrays or genotyping by sequencing (GBS). Genotyping microarrays use specifically designed oligonucleotides to bind matching nucleotide sequences. Afterwards two signals are obtained for each allelic variant of the SNP (e.g., A or T) by laser scans. The intensity and difference of the signals provide information about the genotype. GBS uses read counts to obtain genotypes. It requires high sequencing coverage for the region(s) of interest. The ratio between the numbers of reads coding for each allele reveals the genotype for a specific marker. Both technologies have advantages and disadvantages. Microarrays are cheaper once they are designed, because they can be applied to large numbers of individuals at small costs. In comparison,

GBS is more expensive, because the sequencing costs per individual remain the same. On the other hand, GBS is more flexible and not restricted to the markers on a previously designed chip. At sufficient read depth, the resolution is higher compared to today's microarrays and allows for tri- and tetra-allelic SNPs.

Genotyping tools have to cope with data produced by both technologies. The input consists of one signal value/read count per allele. For each marker the values of all samples are grouped into clusters. Each cluster represents one genotype. It is important to analyze the markers individually, because there are significant differences between the markers. Processing them together, adds an additional layer of noise to the data. There are different ways to transform the input signals  $a$  and  $b$  into values for contrast  $M$  and intensity  $S$  [13-15].  $M$  represents the difference between the two alleles and is used to distinguish between the genotypes.  $S$  stands for the combined signal strength and indicates whether the measurement was successful. Low read counts or intensities indicate non-reliable genotype information. Instead of assigning an uncertain genotype, these values should be called missing. However, they might be imputed in a later stage of the analysis. Diploids have three different genotypes for biallelic SNPs: AA (diplex), AB (monoplex) and BB (nulliplex). The number of possible genotypes increases with the ploidy level. Standard genotyping tools (e.g., `crlmm` [14]) do not support more than three genotypes per SNP. Hence, special tools for polyploids are required. Some of the current tools and their limitations are described below. Besides the ratio between the two alleles, the total intensity  $S$  per sample is of interest.

Markers that exceed a threshold number of low intensity samples should be completely excluded from the analysis. Low signal values indicate problems with the target sequence: In case of microarray genotyping the nucleotide acid sequence fails to bind to the oligonucleotide of the chip and during GBS, the mapping of the target reads to the reference sequence

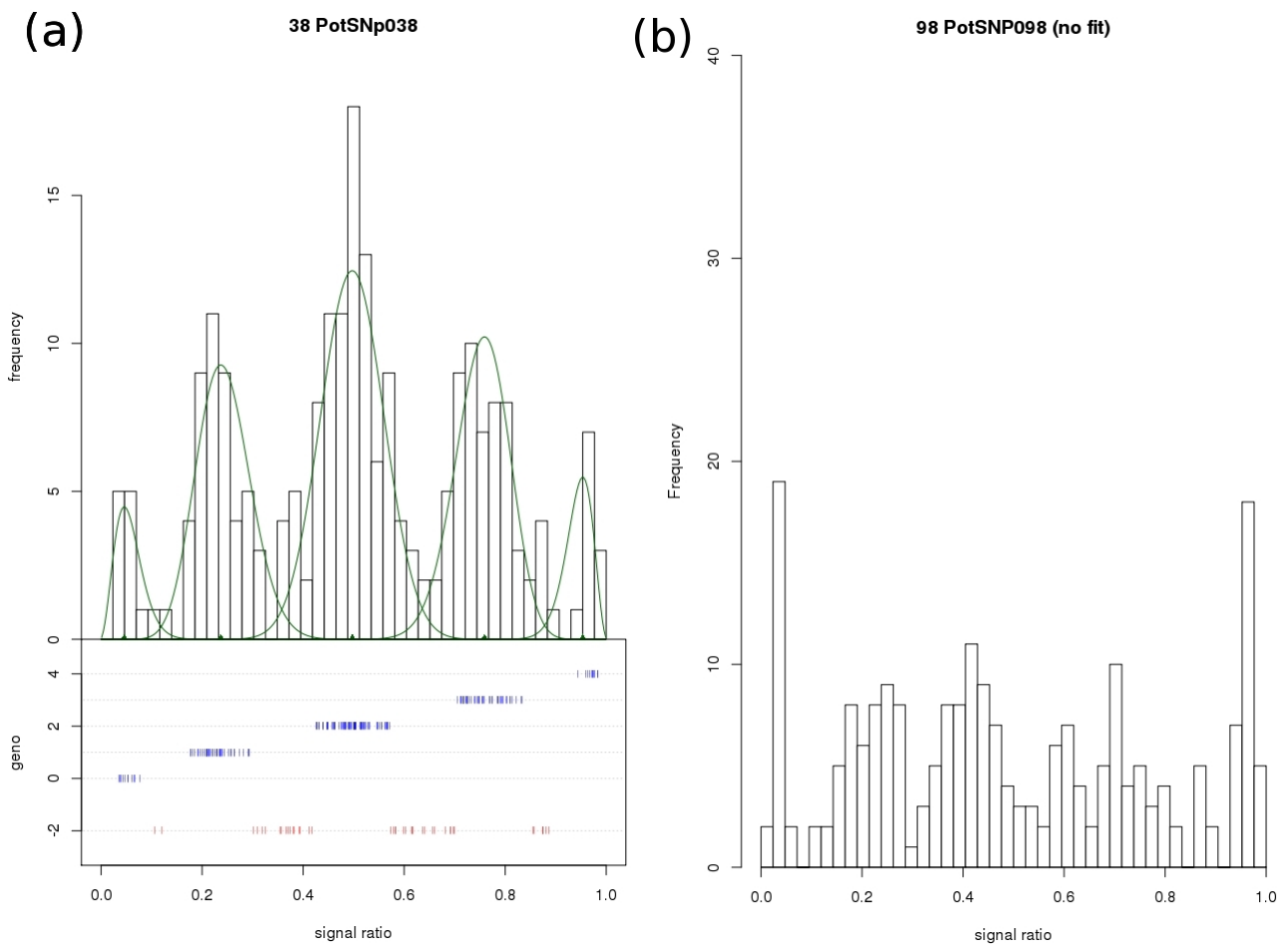
fails. Microarrays produce background signals for both alleles, which can be misinterpreted as hybridization signals. Incorrect genotype calls can influence the downstream analysis and lead to false conclusions. Available solutions differ in the clustering methods that are used to identify the genotypes. In addition, they employ various preprocessing steps to obtain uniform data. The tools described below address the task of polyploid genotype calling with various methods.

#### 2.1.1 FitTetra

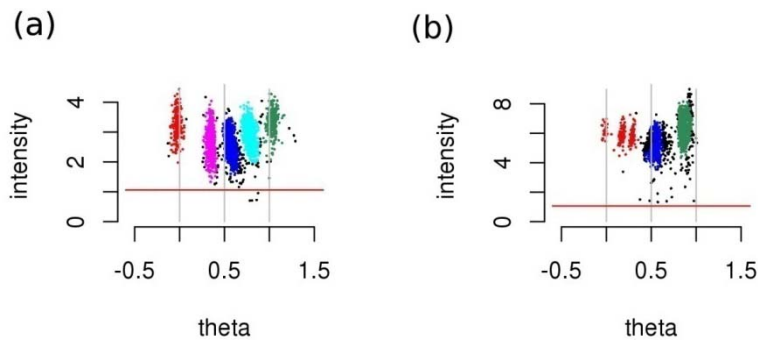
FitTetra [15] is an R [16] package for genotype calling in tetraploids. The supported ploidy levels are two and four. It does not calculate intensity values. Hence, failed samples may not be recognized and are included in the analysis. FitTetra fits eight different mixture models to the data. Afterwards, the results are checked for different criteria in order to exclude some models. The remaining model with the lowest Bayesian Information Criterion (BIC) [17], is chosen for the marker. The multiple-step process comes at the cost of performance. However, it could be shown that the tool is working for an Illumina Golden Gate (Illumina Inc.) data set of tetraploid potato. Fig. 1 shows example plots for two markers. The top part shows the ratios between the two alleles and their frequency in the population as a histogram. The green lines indicate the fitted model. The lower part translates the histogram into genotypes and uncalled samples (-2). Fig. 1a classifies five clusters. Many samples remain uncalled, because their ratios lay between clusters. Fig. 1b is an example of a failed clustering, because none of them odels fitted to the data.

#### 2.1.2 BeadarrayMSV

BeadarrayMSV [18] is another publicly available R package for genotype calling. It aims to identify multisite variants (MSVs), which are tetraploid SNPs in the partially tetraploid Atlantic salmon (*Salmo salar*). However, the genotype calling process is equal to the one in fully tetraploid organisms. BeadarrayMSV requires the data in formats that are



**Fig. 1** FitTetra example plots for two tetraploid markers from the FitTetra dataset.



**Fig. 2** BeadarrayMSV examples from the default datasets of the R package.

unique to Illumina Infinium Bead Arrays (Illumina Inc.). The software package uses k-means clustering to identify the most likely one of seven cluster combinations. These determine the number and position of the clusters. Fig. 2 shows example plots for two markers. Each cluster is shown in a different color, while unassigned samples are black. The red line

indicates the intensity threshold. Fig. 2a shows the classification of a tetraploid marker into five clusters, as expected. Fig. 2b is classified into three clusters, but the left cluster is inconclusive and many samples are identified as noise.

### 2.1.3 Supermassa

Supermassa [19] is a web-based software package

for polyploid genotype calling. To the best of our knowledge, it is the only available tool for ploidy levels higher than four. A graphical Bayesian model estimates the probability of the data given a certain ploidy. Hence, out of a range of possible even-numbered ploidies, the most likely one is returned. The visualization consists of a scatter plot including one line per genotype indicating the average cluster angle. Supermassa's performance is demonstrated by its authors with the before mentioned potato data set and a sugarcane data set (ploidy 6-14) produced with the Sequenom MassARRAY iPLEX® [19]. Fig. 3 shows example plots of the data from fitTetra. In opposite to Fig. 1, both markers are classified as tetraploid and consist of five clusters.

## 2.2 Linkage Mapping

Linkage mapping is the construction of a linkage map. First, genetic markers are grouped into linkage groups that ideally represent chromosomes. Certain crossing schemes allow the estimation of pairwise recombination frequencies between markers. These can be employed to assess the linkage between all markers and to identify linkage groups. The markers within each group are ordered according to their linkage. Mapping functions provide the individual spatial distance between the markers on the linkage map.

Diploid linkage mapping has been successfully applied for decades. The general method is applicable for polyploids, as well. However, the calculation of the recombination frequencies is difficult, because the meiotic mechanism in polyploids is different [20]. For instance, it allows for double reduction, which is when sister alleles are included in the same gamete [21]. Various research groups approached to resolve the issue, but none of the methods have been successfully integrated into a publicly available tool [22-25]

Further, there is a discussion about the underlying assumptions and limitations of the methods by the authors of the methods [25, 26]. A current workaround is the usage of single-dose markers, which behave like diploids [27].

To our knowledge, two tools have been developed for tetraploid linkage mapping. Both of them are limited to simple models and data sets. None of them is applicable for today's SNP data sets. However, their development shows a pending need for polyploid linkage mappers for more than ten years.

### 2.2.1 Polylink

Polylink [28] supports two-point linkage analysis in even-numbered polyploids. It aims to estimate the recombination fraction from observed data. However, it does not calculate all pairwise combinations of markers. The algorithm is based on a random chromosome pairing model. Polylink uses a simplified

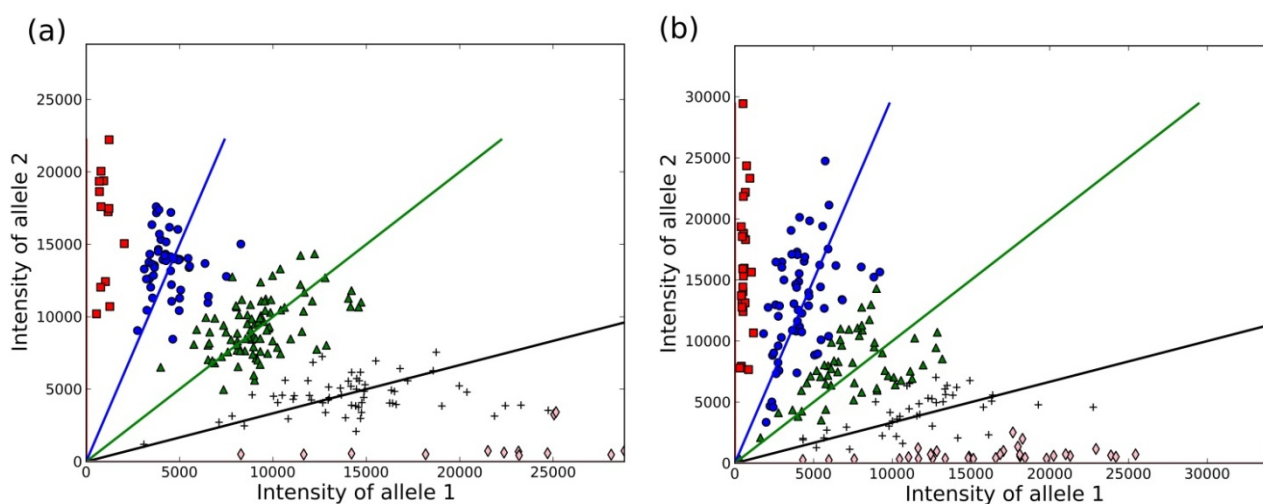


Fig. 3 Supermassa examples from the FitTetra dataset (SNPs 38 and 98).

model, which ignores double reductions. In addition, it provides a simulation facility. The simulation requires the parental phase, the recombination fraction and the number of offspring to be known.

#### 2.2.2 TetraploidMap (for Windows)

TetraploidMap [29] is a tool for linkage mapping of single- and double-dose, multiallelic, dominant and autotetraploid markers. It was extended in 2007 to TetraploidMap for Windows [30]. The updated version provides QTL interval mapping and a graphical user interface. Nevertheless, the limitation to the single- and double-dose markers persists.

### 2.3 Haplotype Phasing

Haplotypes are clusters of SNPs, inherited together. Thus, they provide further information about recombination events and outperform single marker analysis [31]. There are multiple approaches to address this task in diploids. All of them are computationally expensive. Most of them are based on hidden markov models (HMMs). In these approaches, the clusters of the HMMs represent common haplotypes. The number of possible combinations of the clusters at each marker causes the computation costs of the methods. For polyploids, the number of clusters is usually higher, because there are more variation possibilities. Hence, the runtime increases with the level of ploidy. Large data sets for polyploids require fast heuristics to resolve the haplotypes.

#### 2.3.1 Polyhap

Polyhap [32] is a Java implementation of a hidden Markov model (HMM) based method to infer haplotypes. The genotypes of all individuals (observations) are used to infer the ancestral haplotypes (hidden states). The latter ones are used to predict the individual haplotypes afterwards. The method allows different individuals to have different ploidy levels. Polyhap is applied to both simulated and real data sets and its performance is compared to other tools [32].

#### 2.3.2 SATlotyper

SATlotyper [33] is a Java implementation of a Boolean satisfiability problem (SAT) solver based haplotype inference method. It is publicly available from the GABI Primary Database website [34]. The tool aims to minimize the number of haplotypes that is required to explain the input data (SNPs). This strategy is called Haplotype Inference by Pure Parsimony (HIPP) [35]. SAT can be used to describe the problem and thus, SAT solving approaches can be applied to haplotype phasing [36]. SATlotyper extends this approach to polyploids. It is applied to simulated and real data sets. A bootstrapping procedure compares different solution and provides probabilities for them. The application of SATlotyper to data sets with a size of 25 or more is computationally not feasible. In addition, the memory usage exceeds the limits of the computers used in the analysis [32]. The runtime increases exponentially with the number of markers [33].

### 2.4 Miscellaneous

Besides the three main types genotype callers, linkage mappers and haplotype phasers, there are other tools for polyploids.

#### 2.4.1 PedigreeSim

In order to validate linkage mapping tools, it is required to have data sets where the linkage map is known. Real data sets of that kind are rarely available for polyploids, but they can be simulated. The simulation process has to take the ploidy level into account. For instance, in tetraploids there can be either two bivalents or one quadrivalent. This does not have to be equal for all chromosomes or individuals. In a hexaploid, there could be three bivalents, one bivalent and one quadrivalent or one hexavalent [22]. The number of possible combinations increases with the level of ploidy.

PedigreeSim [12] is a publicly available Java implementation for such a simulation of di and tetraploids. It takes a population structure, parental

genotypes and a genetic map as input and simulates offspring genotypes. Further parameters can be specified (e.g., the bivalent to quadrivalent ratio and preferential pairing probabilities).

#### 2.4.2 PolySegratio

PolySegratio [37] is an R package to analyze the segregation ratios of autopolyploids. It is not limited to a certain level of ploidy. However, it gives a warning, if odd-numbered ploidies are entered. In addition, PolySegratio provides functions to simulate data sets. The numbers of individuals and markers, as well as the ploidy levels and genotype dosages can be specified. The output consists of data objects, as well as data visualizations. The limitations are similar to polylink and tetraploid map, because only dominant markers are supported.

#### 2.4.3 PolyCat

PolyCat [38] is a sequencing read mapping pipeline and genome categorization tool for allopolyploids. It distinguishes between allele-SNPs and homeo-SNPs, which are polymorphisms within and between subgenomes, respectively. The latter ones are used to identify the two subgenomes of the allopolyploid. The method uses the Genomic Short-read Nucleotide Alignment Program (GSNAP) for SNP-tolerant mapping of the reads to Ref. [39]. PolyCat is limited to allopolyploids and requires one homeo-SNP within each read. These restrictions require many SNPs between the subgenomes and a long sequencing read length. However, the approach has been successfully applied to cotton (*Gossypium hirsutum*,  $2n = 4x = 52$ ),

rapeseed (*Brassica napus*,  $2n = 4x = 38$ ) and wheat (*Triticum aestivum*,  $2n = 6x = 42$ ) [38].

### 3. Discussion

The previous sections show a variety of tools for polyploids. However, we showed that all of them have limitations to certain ploidy levels or in their performance (Table 1). For instance, haplotyping supports all levels of ploidy, but the computing time for a small tetraploid data set is more than a day [32]. Hence, there is still the need for methods with increased performance in order to allow for larger data sets, higher ploidies or multiple runs.

Moreover, there are fields where no tools are available for polyploids (e.g., genome assembly). Assembling sequenced reads of DNA into a genome is already difficult for diploids. The challenge is to distinguish between sequencing errors and genetic variants. For instance, in octoploids the support can be as low as 1:8 for a variant and ABBBBBBB has to be differentiated from AABBBBBB and BBBBBBBB. High sequencing coverage or signal intensities are required to obtain unambiguous results in that case. The problem remains when a reference genome is available and read mapping is applied. Prospectively developed tools for assembly and read mapping should take polyploids into account.

Genotype calling for tetraploids is now well established [40]. Nevertheless, low intensity samples do not have lower probabilities or are excluded from the analysis. For higher ploidy levels only Supermassa

**Table 1** Overview of the tools and their supported ploidy levels.

Tool	Year	Ploidy levels	Auto-/allo-ploids	Type
FitTetra [15]	2011	2, 4	Both	Genotype calling
BeadarrayMSV [16]	2011	2, 4	Both	Genotype calling
Supermassa [19]	2012	Any even-numbered	Both	Genotype calling
PedigreeSim [13]	2012	2, 4	Both	Genotype simulation
Polylink [28]	2001	2, 4, 6, 8, 10	Autoploids	Linkage mapping
TetraploidMap for Windows [30]	2007	4	Autoploids	Linkage mapping
Polyhap [32]	2008	Any level	Both	Haplotype phasing
SATlotyper [33]	2008	Any level	Both	Haplotype phasing
PolySegratio [37]	2008	Any level	Autoploids	Segregation ratio analysis
PolyCat [38]	2013	Any level	Allopolyploids	Subgenome categorization

exists. It is limited to a web interface. Thus, it cannot easily be integrated into an automated pipeline. Moreover, it has been shown to overestimate the ploidy level, because it does allow for clusters with low support [2]. Besides the two outdated implementations and the three before mentioned theoretical models of linkage in polyploids, there are no functional linkage mapping tools available. Usage of single-dose markers is a limitation that needs to be overcome.

#### 4. Conclusions

Many of the polyploid crops are commercially important. While more and more tools are available for diploids, the research of polyploid species lags behind. However, prices for genotyping are decreasing and the toolbox for polyploids will certainly grow, because certain polyploids are main players in the global food market (e.g., wheat and potato). Some tools reviewed in this article were designed especially for tetraploids. In the future, they might be extended to higher ploidy, as well. As soon as high quality data sets are available for higher ploidies, the research in this field will be boosted. In addition to the real data sets, data simulation methods will provide data to validate the developed tools. In summary, polyploids are essentials for the world's food supply and plant breeding will strongly improve in this field.

#### Acknowledgment

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 289974.

#### References

- [1] Comai, L. 2005. "The Advantages and Disadvantages of Being Polyploid, Nat." *Rev. Gen.* 6: 836-46.
- [2] Garcia, A. A. F., Mollinari, M., Marconi, T. G., Serang, O. R., Silva, R. R., Vieira, M. L. C., Vicentini, R., Costa, E. A., Mancini, M. C., Garcia, M. O. S., Pastina, M. M., Gazaffi, R., Martins, E. R. F., Dahmer, N., Sforça, D. A., Silva, C. B. C., Bundock, P., Henry, R. J., Souza, G. M., vanSluys, M. A., Landell, M. G. A., Carneiro, M. S., Vincentz, M. A. G., Pinto, L. R., Vencovsky, R., and Souza, A. P. 2013. "SNP Genotyping Allows an in-Depth Characterisation of the Genome of Sugarcane and Other Complex Autopolyploids." *Sci. Rep.* 3: 3399.
- [3] Song, C., Liu, S., Xiao, J., He, W., Zhou, Y., Qin, Q., Zhang, C., and Liu, Y. 2012. "Polyploid Organisms." *Sci. China Life Sci.* 55: 301-11.
- [4] Pasquier, L. D. 2009. "The Fate of Duplicated Immunity Genes in the Dodecaploid *Xenopus ruwenzoriensis*." *Front. Biosc.* 14: 177-91.
- [5] Acquah, G. 2007. *Principles of Plant Breeding and Genetics*. Malden, MA USA: Blackwell Publishing.
- [6] Wu, J. H. A., Ferguson, R., Murray, B. G., Jia, Y., Datson, P. M., and Zhang, J. 2012. "Induced Polyploidy Dramatically Increases the Size and Alters the Shape of Fruit in *Actinidia chinensis*." *Ann. Bot.* 109: 169-79.
- [7] Ramsey, J., and Schemske, D. W. 1998. "Schemske, Pathways, Mechanisms and Rates of Polyploid Formation in Flowering Plants." *Annu. Rev. Ecol. Syst.* 29: 467-501.
- [8] Compton, M. E., Gray, D. J., and Elmstrom, G. W. 1996. "Identification of Tetraploid Regenerants from Cotyledons of Diploid Watermelon Cultured *in Vitro*." *Euphytica* 87: 165-72.
- [9] Mammadov, J., Aggarwal, R., Buyyarapu, R., and Kumpatla, S. 2012. "SNP Markers and Their Impact on Plant Breeding." *International Journal of Plant Genomics* 2012: 1-11.
- [10] Hodgkinson, A., and Eyre-Walker, A. 2010. "Human Triallelic Sites: Evidence for a New Mutational Mechanism?" *Genetics* 184: 233-41.
- [11] Lynch, M., and Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sunderland: Sinauer Associates.
- [12] Xu, S. 2012. *Principles of Statistical Genomics*. NY: Springer Science & Business Media.
- [13] Voorrips, R. E., and Maliepaard, C. A. 2012. "The Simulation of Meiosis in Diploid and Tetraploid Organisms Using Various Genetic Models." *BMC Bioinf.* 13: 248.
- [14] Carvalho, B., Bengtsson, H., Speed, T. P., and Irizarry, R. A. 2007. "Exploration, Normalization and Genotype Calls of High-Density Oligonucleotide SNP Array Data." *Biostat.* 8: 485-99.
- [15] Voorrips, R. E., Gort, G., and Vosman, B. 2011. "Genotype Calling in Tetraploid Species from Bi-Allelic Marker Data Using Mixture Models." *BMC Bioinf.* 12: 172.
- [16] Team, R. D. C. 2011. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.



- [17] Schwarz, G. 1978. "Estimating the Dimension of a Model." *Ann. Statist.* 6: 461-4.
- [18] Gidskehaug, L., Kent, M., Hayes, B. J., and Lien, S. 2011. "Genotype Calling and Mapping of Multisite Variants Using an Atlantic Salmon iSelect SNP Array." *Bioinf.* 27 (3): 303-10.
- [19] Serang, O., Mollinari, M., and Garcia, A. A. F. 2012. "Efficient Exact Maximum a Posteriori Computation for Bayesian SNP Genotyping in Polyploids." *PLoS ONE*. 7 (2): e30906.
- [20] Fisher, R. A. 1947. "The Theory of Linkage in Polysomic Inheritance." *Soc. Lond. B.* 233: 55-87.
- [21] Mather, K. 1936. "Segregation and Linkage in Autotetraploids." *Journ. of Genetics* 32: 287-314.
- [22] Rehmsmeier, M. 2013. "A Computational Approach to Developing Mathematical Models of Polyploid Meiosis." *Genetics* 193: 1083-94.
- [23] Wu, R., Gallo-Meagher, M., Littell, R. C., and Zeng, Z. B. 2001. "A General Polyploid Model for Analyzing Gene Segregation in Outcrossing Tetraploid Species." *Genetics* 159: 869-82.
- [24] Luo, Z. W., Zhang, R. M., and Kearsy, M. J. 2004. "Theoretical Basis for Genetic Linkage Analysis in Autotetraploid Species." *Proc. Nat. Acad. of Sci. of the U.S.A.* 101: 7040-5.
- [25] Wu, R., and Ma, C.-X. 2005. "A General Framework for Statistical Linkage Analysis in Multivalent Tetraploids." *Genetics* 170: 899-907.
- [26] Luo, Z. W. 2005. "Commentary on Wu and Ma." *Genetics* 171 (4): 2149-50.
- [27] Bertioli, D. J., Ozias-Akins, P., Chu, Y., Dantas, K. M., Santos, S. P., Gouvea, E., Guimarães, P. M., Leal-Bertioli, S. C. M., Knapp, S. J., and Moretzsohn, M. C. 2014. "The Use of SNP Markers for Linkage Mapping in Diploid and Tetraploid Peanuts." *G3 (Bethesda)* 4: 89-96.
- [28] He, Y., Xu, X., Tobutt, K. R., and Ridout, M. S. 2001. "Polylink: To Support Two-Point Linkage Analysis in Autotetraploids." *Bioinf.* 17: 740-1.
- [29] Hackett, C. A., and Luo, Z. W. "TetraploidMap: Construction of a Linkage Map in Autotetraploid Species." *J. Hered.* 94: 358-9.
- [30] Hackett, C. A., Milne, I., Bradshaw, J. E., and Luo, Z. "TetraploidMap for Windows: Linkage Map Construction and QTL Mapping in Autotetraploid Species." *J. Hered.* 98: 727-9.
- [31] Mailund, T., Besenbacher, S., and Schierup, M. H. 2006. "Whole Genome Association Mapping by Incompatibilities and Local Perfect Phylogenies." *BMC Bioinf.* 7: 454.
- [32] Su, S.-Y., White, J., Balding, D. J., and Coin, L. J. M. 2008. "Inference of Haplotypic Phase and Missing Genotypes in Polyploid Organisms and Variable Copy Number Genomic Regions." *BMC Bioinf.* 9: 513.
- [33] Neigenfind, J., Gyetvai, G., Basekow, R., Diehl, S., Achenbach, U., Gebhardt, C., Selbig, J., and Kersten, B. 2008. "Haplotype Inference from Unphased SNP Data in Heterozygous Polyploids Based on SAT." *BMC Genomics*. 9: 356.
- [34] Usadel, B., Schwacke, R., Nagel, A., and Kersten, B. 2012. "GabiPD—The GABI Primary Database Integrates Plant Proteomic Data with Gene-Centric Information." *Front. Plant Sci.* 3: 154.
- [35] Gusfield, D. 2003. "Haplotype Inference by Pure Parsimony." *Lecture Notes in Computer Science* 2676: 144-55.
- [36] Lynce, I., and Marques-Silva, J. 2006. "Efficient Haplotype Inference with Boolean Satisfiability." In *Proceedings of the 21st National Conference on Artificial Intelligence*, 104-9.
- [37] Baker, P. 2008. "PolySegratio: An R Library for Autopolyploid Segregation Analysis." PolySegratio Library. Accessed January 9, 2008. <http://cran.r-project.org/web/packages/polySegratio/vignettes/polySegratio-overview.pdf>.
- [38] Page, J. T., Gingle, A. R., and Udall, J. A. 2013. "PolyCat: A Resource for Genome Categorization of Sequencing Reads From Allopolyploid Organisms." *G3 (Bethesda)* 3: 517-25.
- [39] Wu, T. D., and Nacu, S. 2010. "Fast and SNP-Tolerant Detection of Complex Variants and Splicing in Short Reads." *Bioinf.* 26: 873-81.
- [40] Uitdewilligen, J. G. A. M. L., Wolters, A. M. A., D'hoop, B. B., Borm, T. J. A., Visser, R. G. F., and van Eck, H. J. 2013. "A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato." *PLoS ONE* 8 (5): e62355.