ApplicationNote

affymetrix
Biology for a better world

Allopolyploid Genotyping Algorithm on Affymetrix Axiom® Arrays

Ali Pirani, Hong Gao, Laurent Bellon, and Teresa A. Webster

Abstract

Many plant and animal species present unique challenges to accurate genotyping due to complexity associated with polyploid genomes. In allopolyploid species, SNPs segregating in 1 subgenome tend to exhibit particular clustering behaviors due to the allele-dosage contribution of the alternate subgenomes. AxiomGT1 is an automated clustering and genotype assignment algorithm used by Axiom Genotyping Console™ Software, and it features adaptive, dynamic clustering that employs statistical methods for accurate genotyping of allopolyploid species.

The Affymetrix® Software employs a statistical approach called "FitAllo™ cluster prediction tool" to produce SNP-specific priors that are then applied by AxiomGT1, the genotyping algorithm in Axiom Genotyping Console Software. The SNP-specific priors provide the genotyping software with hints to accurately ascertain specific SNP cluster patterns in order to derive the center locations of each genotype cluster and automatically assign the correct genotype. We applied this process to bread wheat varietal (allohexaploid) genotype data generated on Axiom custom genotyping arrays and demonstrated accurate genotype calls in allopolyploids. Genotyping Console Software with SNP-specific priors generated using FitAllo cluster prediction tool is widely applicable in genotyping allopolyploid plant and animal species.

Background

In allopolyploid species, SNPs tend to segregate in 1 subgenome. As a result, the SNPs demonstrate 3 major genotype clustering patterns: the diploid-like pattern where sub-genomes' effects counteract each other (denoted as AB/(AA)n/(BB)n) or 1 of 2 compressed, shifted patterns (denoted as AB/(BB)n or AB/(AA)n). See Figure 1 for examples of these 3 cluster patterns. For the latter 2 patterns, other genotyping platforms and technologies are likely to either not call the genotypes or to miscall the shifted homozygous cluster as heterozygotes and indicate that the minor homozygous genotypes are missing. This is more likely to occur in inbred populations when there are few or no heterozygous samples, as shown in Figure 1B and 1C.

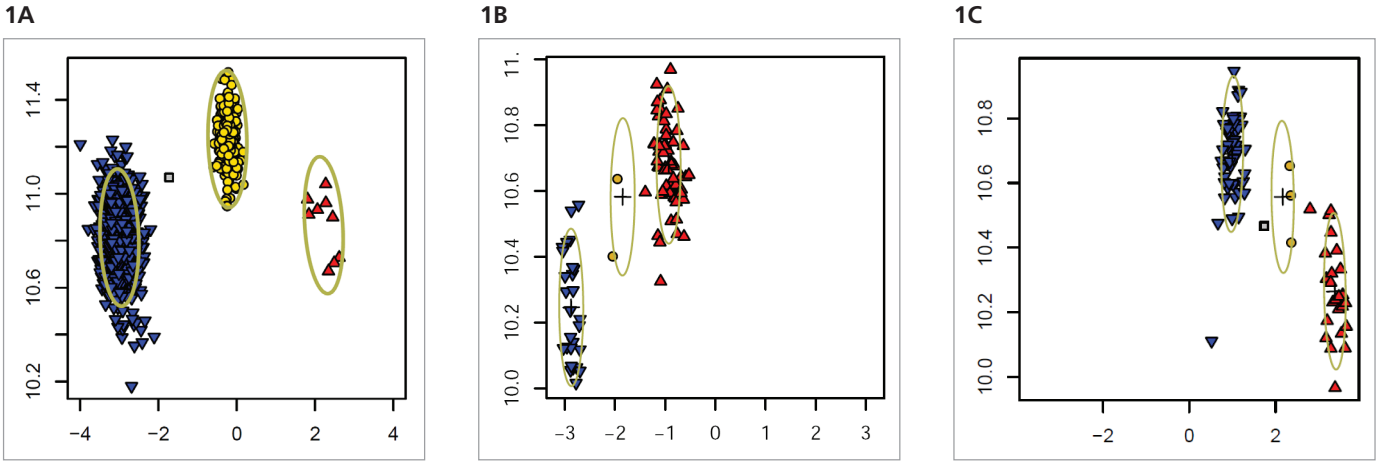

Figure 1. Allopolyploid genotype cluster plots. (A) The diploid-like pattern has the heterozygous cluster centered in contrast space. (B) Allopolyploid cluster plot with compressed clusters shifted to the left in contrast space with the pattern AB/(BB)n. (C) Allopolyploid cluster plot with compressed clusters shifted to the right with the pattern AB/(AA)n.

Consequently, manual genotype checking and correction is required. This is labor intensive and time consuming, as it can require a whole day for every 1,000 SNPs that require checking. This is a considerable analysis burden, even for a low-density SNP panel.

FitAllo™ cluster prediction tool enables the creation of SNP-specific priors that are then used by Axiom Genotyping Console™ Software to correctly assign genotype calls to these SNPs. Axiom® Genotyping Solution can accurately identify the heterozygous samples for such SNPs from plants and animals with allopolyploid genomes. This analysis can be completed in 1 hour, regardless of the number of SNPs in the genotyping panel.

The FitAllo cluster prediction tool algorithm chooses the best fit for each cluster pattern among the multiple pre-specified allo-generic priors via the Baum-Welch algorithm, as shown in Figure 2.

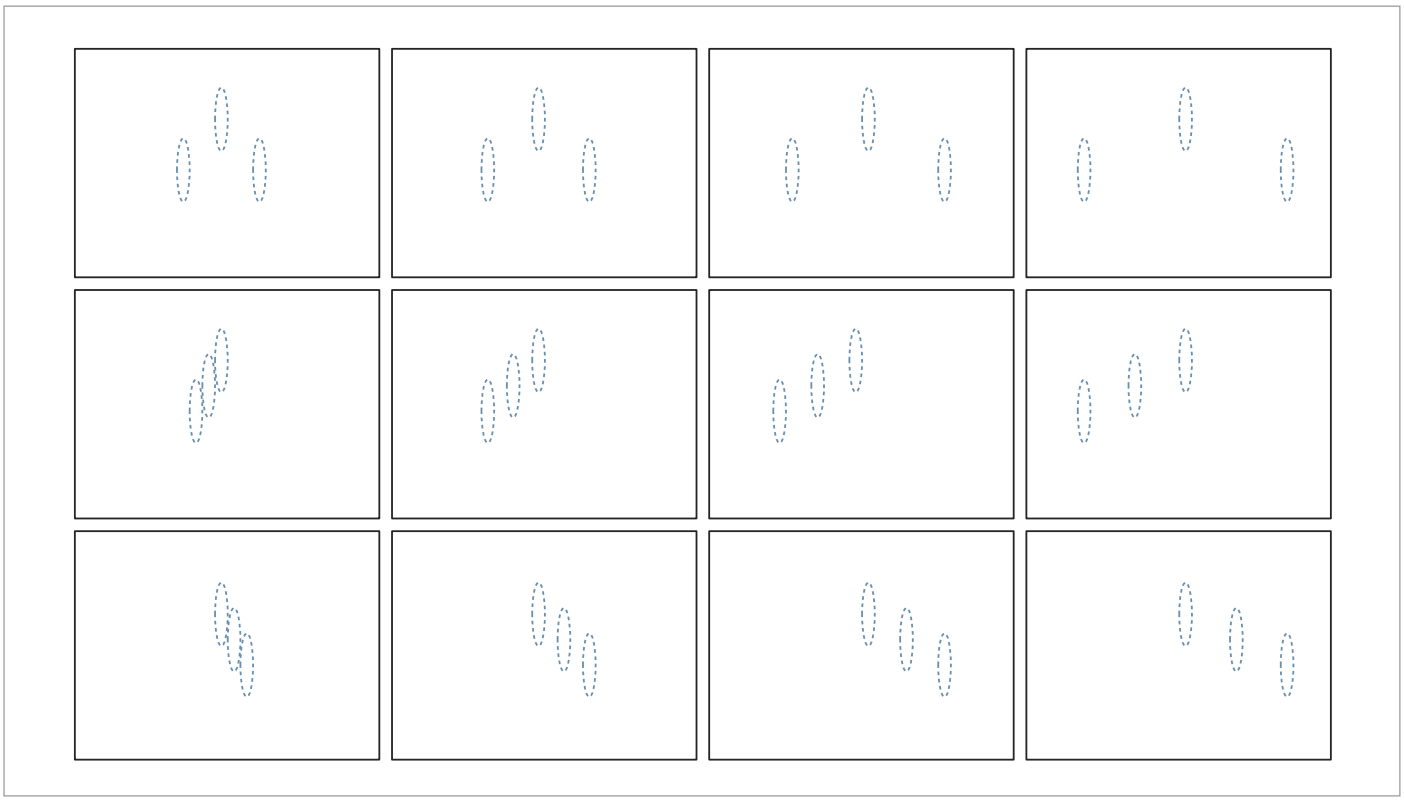

Figure 2. Prior models that FitAllo cluster prediction tool assigns to allopolyploid genotype cluster plots.

After ascertaining specific SNP patterns [AB/(AA)n/(BB)n or AB/(AA)n or AB/(BB)n], the center locations of each genotype cluster are located, which can then be used as SNP-specific priors for the AxiomGT1 algorithm. This becomes especially useful when populations are inbred and have decreased heterozygosity for some of the genotyped SNPs, which will cause those SNPs to exhibit 2 homozygous clusters. With only 2 clusters present, the prior models aid in ascertaining the genotype call that should be assigned to each of the clusters. It is also useful for very compressed clusters that do not have clear separation from one another.

FitAllo cluster prediction tool correctly fits each genotype into 1 of the models shown above and generates SNP-specific priors. AxiomGT1 uses these priors to correctly call off-centered clusters and improve the overall genotyping accuracy. See Figure 3 for the complete view of the automated data analysis workflow.

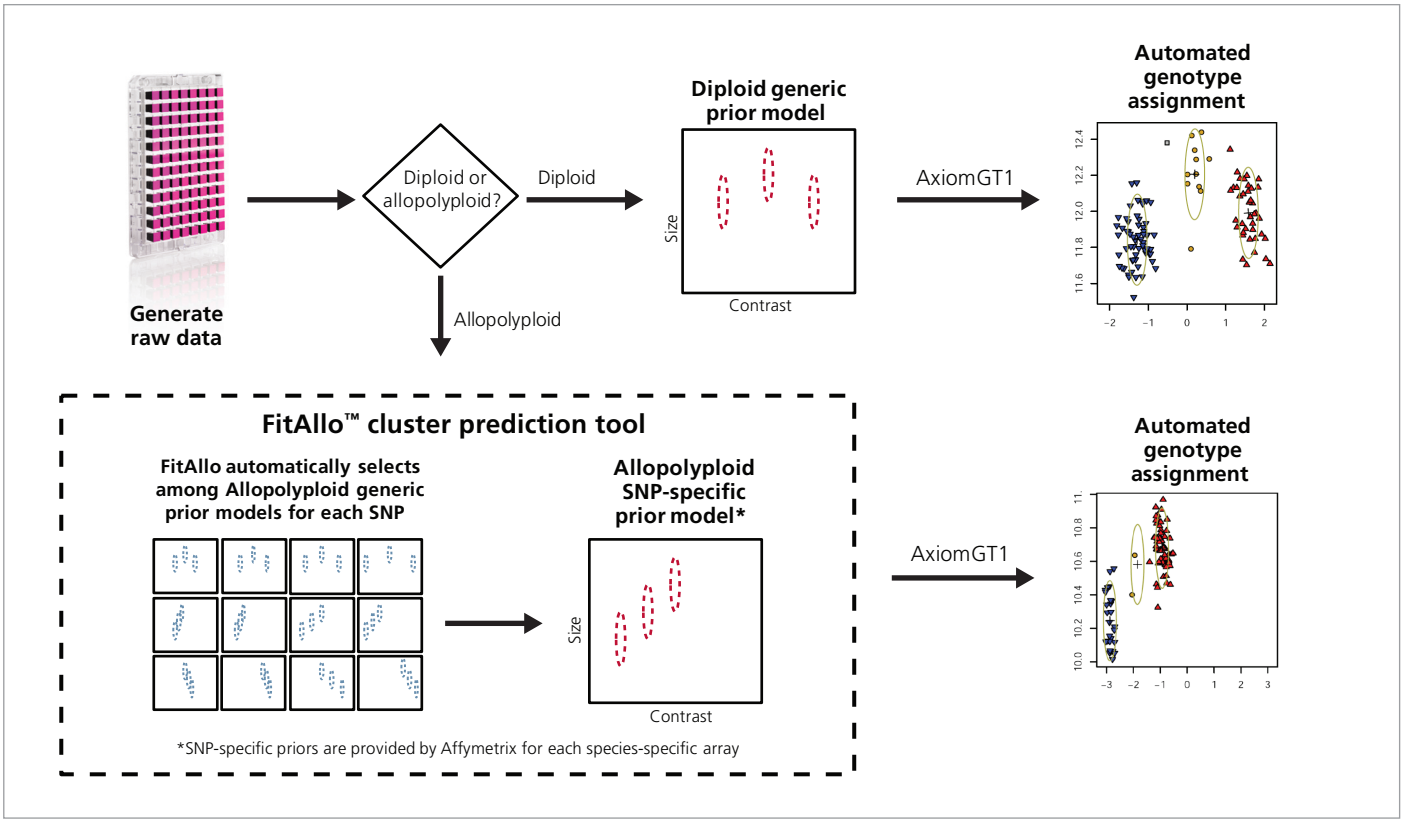

Figure 3. Genotyping data analysis process with Axiom Genotyping Console™ Software for both diploid and allopolyploid species.

Case study: Application to bread wheat

We applied FitAllo™ cluster prediction tool to bread wheat varietals (allohexaploid) genotype data from customized Axiom myDesign™ Genotyping Arrays in collaboration with University of Bristol, UK. This dataset contains 11,603 SNPs genotyped across 89 samples.

Figure 4 shows 2 examples of SNPs with decreased heterozygosity correctly called by AxiomGT1. The usage of SNP-specific prior models generated by FitAllo cluster prediction tool enables accurate genotype assignment of both homozygotes and heterozygotes.

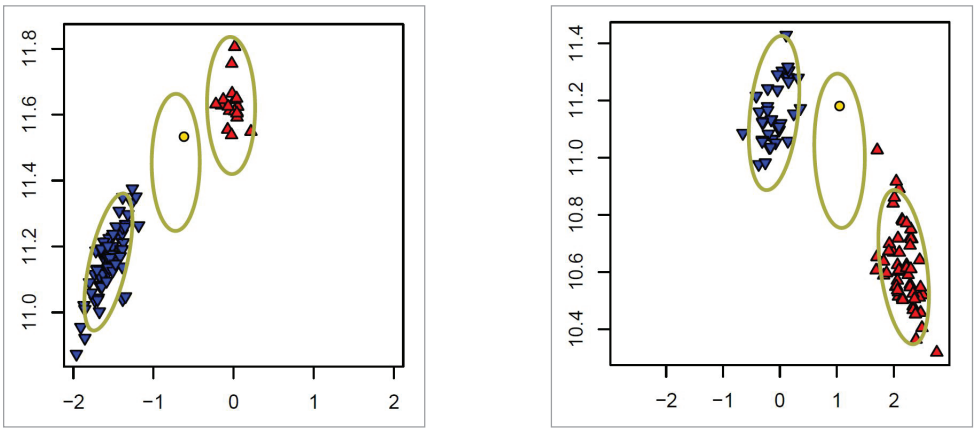

Figure 4. Two examples of wheat genotypes with low representation of heterozygotes. The SNP on the left is accurately called as AA-AB-BB. The SNP on the right is accurately called as BB-AB-AA. The usage of SNP-specific prior models by AxiomGT1 enables accurate genotype assignment of both homozygotes and heterozygotes.

After the genotype calling was complete, we utilized SNPolisher analysis and visualization package to post-process the genotyping results. SNPolisher analysis and visualization package is an R package offered by Affymetrix that is integrated with Genotyping Console™ Software. It classifies the SNPs into categories and indicates the number of SNPs that fall into each of the categories. All of the categories and their descriptions are shown in Figure 5.

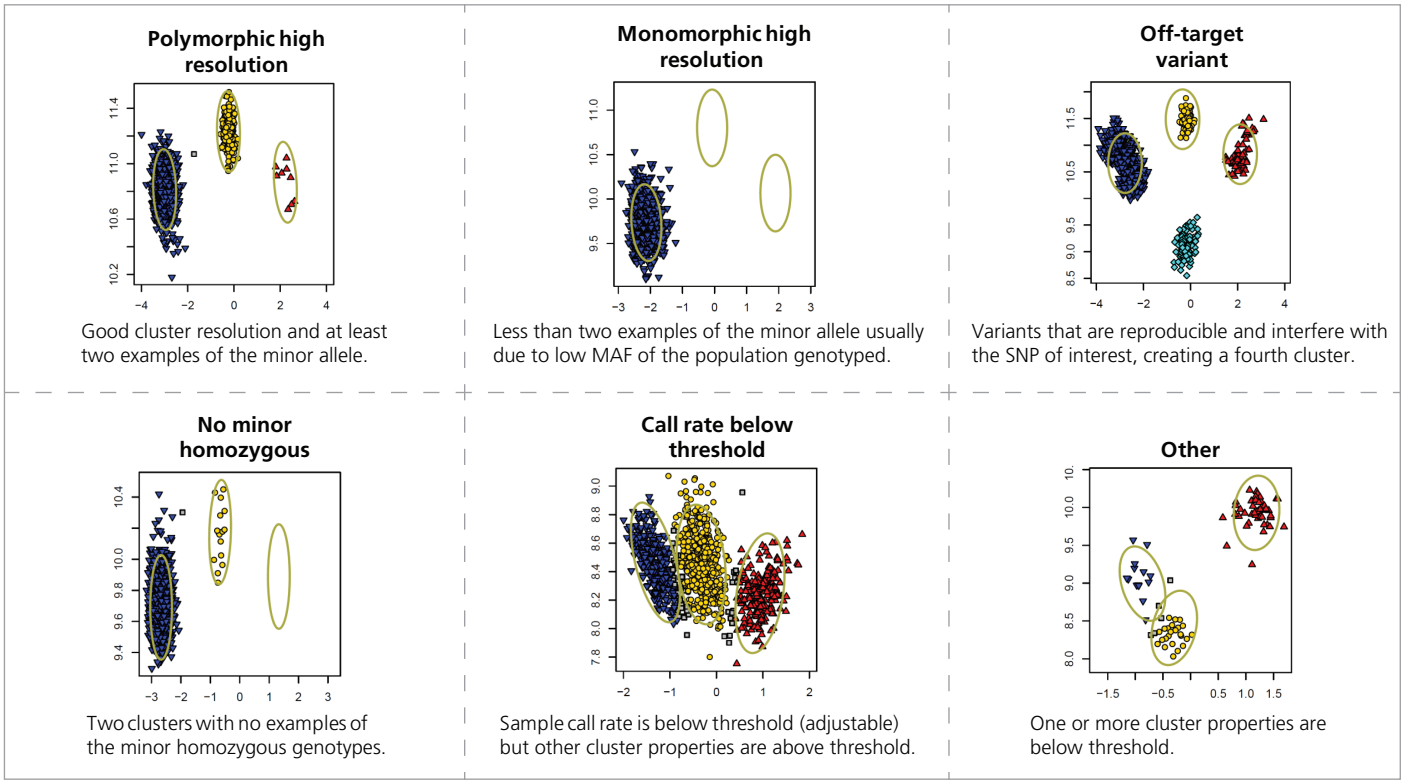

Figure 5. SNPolisher analysis and visualization package genotype classifications and their descriptions.

Ideally, all SNPs will fall into the "polymorphic high resolution" category, which is characterized by good cluster resolution and at least 2 examples of the minor allele. After running FitAllo™ cluster prediction tool to process the bread wheat data, the number of SNPs that fell into the polymorphic high resolution category increased by 15%. A majority of these SNPs were previously classified as "no minor homozygous." This impressive result shows FitAllo cluster prediction tool's utility for ensuring genotyping accuracy in allopolyploids. The clear technical benefit and the days of time savings make Genotyping Console Software with FitAllo cluster prediction tool the ideal solution for high-density genotyping data analysis.

References

1. Voorrips R. E., Gort G., Vosman B. Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinformatics* **12**:172 (2011).

Affymetrix, Inc. Tel: +1-888-362-2447 ■ **Affymetrix UK Ltd.** Tel: +44-(0)-1628-552550 ■ **Affymetrix Japan K.K.** Tel: +81-(0)3-6430-4020
Panomics Solutions Tel: +1-877-726-6642 panomics.affymetrix.com ■ **USB Products** Tel: +1-800-321-9322 usb.affymetrix.com

www.affymetrix.com Please visit our website for international distributor contact information.
For Research Use Only. Not for use in diagnostic procedures.

P/N DNA01921 Rev. 2