

GRAFGEN : A program to design precision graphical genotypes

Bertrand Servin and Frédéric Hospital

Station de Génétique Végétale INRA / CNRS / UPS / INAP-G 91190 Gif-sur-Yvette, France

ABSTRACT

Summary: GRAFGEN is a tool for the analysis of complex breeding schemes with molecular markers. It produces numerical output of probabilities of allelic transmission through complex pedigrees for detailed knowledge of the genomic composition of a population and various graphical representations of the results. In particular, GRAFGEN designs *precision graphical genotypes* which extend the concept of graphical genotypes by interpolating the genotypes between markers while taking into account all possible recombinations given markers map and observed genotypes and pedigree data.

Availability: GRAFGEN is a free software available at <http://moulon.inra.fr/~servin/grafgen>.

Contact: servin@moulon.inra.fr

In most plants and animals, the genotypes of individuals at molecular marker loci located on genetic maps can be assessed, leading to a discrete knowledge of the genomic composition of these individuals. Young and Tanksley (1989) introduced the concept of *graphical genotypes* in order to estimate and visualize the genomic composition of individuals between markers. Graphical genotypes is an helpful tool, for example, to screen populations for individuals carrying desired genotypes at genomic regions of interest, and/or spot favorable recombination events in marker-assisted selection programs. However, the method proposed by Young and Tanksley to build graphical genotypes is an approximation and does not use all the information available.

First, for schemes lasting more than one generation, the accumulation of crossovers over time can not always be ignored. Taking the number of meioses into account gives a better estimate of the genomic composition of individuals.

Second, when available, taking account of the genotypes of the ancestors of the studied individuals in a pedigree can help to determine the most likely set of recombination events that led to the observed data. This set is not necessarily the one implying the fewest crossovers.

Third, taking into account the genotypes at more than two markers flanking the region of interest in a multilocus analysis can help to assess more precisely the set of recombination events.

Finally, more complex breeding schemes than F_2 or backcrosses are commonly used nowadays such as backcross followed by selfing to fix introgressions, or alternation of random mating and selfing to produce Highly Recombinant Inbred Lines. The combination of mating systems used to produce the population strongly affect the rules of allelic transmissions from parents to their offspring, and hence must not be ignored.

GRAFGEN was written in order to take into account all these informations for the estimation of the genomic composition of individuals, and to draw accordingly *precision graphical genotypes*. The principle is to compute the frequencies of all possible genotypes at equally spaced points (virtual loci) on the genome, given all pedigree information. From the results of these computations, GRAFGEN produces the *precision graphical genotype* of an individual. Different representations are possible (see Figure 1).

The computation basis of GRAFGEN is the analytic equations derived by Hospital *et al.* (1996). These equations are implemented in the MDM program (Servin *et al.* , 2002) for numerical computations. These computations allow to estimate the genomic composition of an individual given (if available): i) its genotype at markers, ii) the genotypes of its ancestors, and iii) the breeding scheme from which it is derived. This breeding scheme is any combination of mating

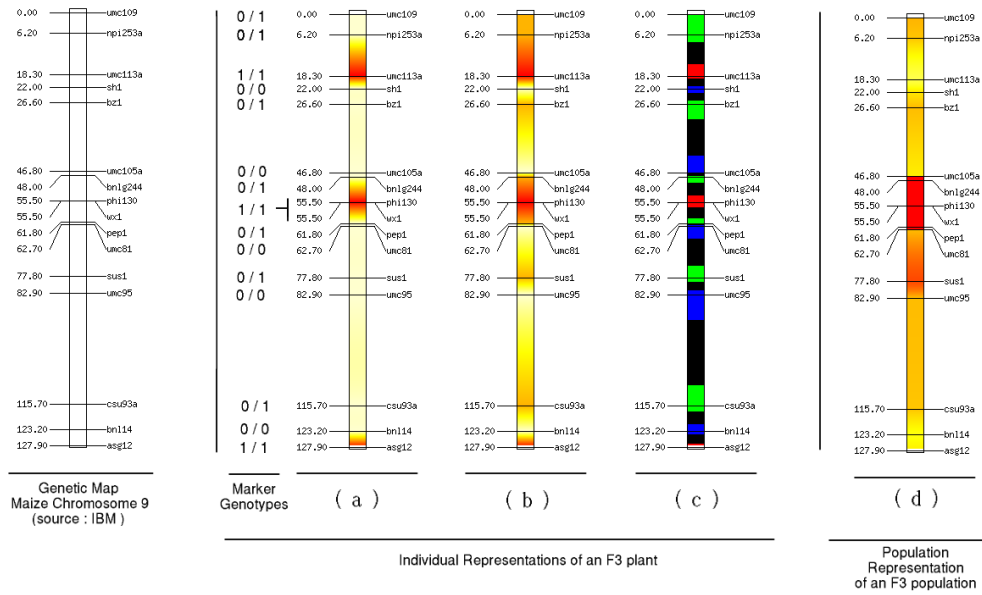


Figure 1: Possible representations of *Precision Graphical Genotypes* - Example of an F_3 population with two alleles segregating (noted 0 and 1). Grafgen represents for each individual either : (a) the probability of being of a given genotype (here the homozygote 1/1), or (b) the expected dose of a particular allele (here 1), or (c) the zones where the probabilities of given genotypes exceed a given threshold (here, the zones of probability > 0.8 are green for the heterozygote 0/1, red for the homozygote 1/1, and blue for homozygote 0/0) ; Grafgen can also represent a synthetic “genotype” for the whole population, according to the mean allele frequency in the population (d).

systems (hybrid mating, selfing, full-sib mating, random mating or doubled haploids). Marker genotypes can be either completely known (including coupling/repulsion phase), or partially known (*e.g.* for dominant markers), or completely unknown (missing data). Hence, the *precision graphical genotypes* produced by GRAFGEN take into account all recombination configurations consistent with the pedigree data while weighting them according to their probabilities of occurrence.

GRAFGEN is a program that produces image files on output (in jpeg or PNG format). It also produces a simple text file containing the results of its computations. GRAFGEN is written in ANSI C using the GD Library (<http://www.boutell.com/gd>). The source code and binaries for Linux and Windows are provided (see Availability). GRAFGEN is a free software published under the GNU General Public License (GPL).

References

- Hospital F, Dillmann C, and Melchinger AE, 1996. A general algorithm to compute multilocus genotype frequencies under various mating systems. *Comput Appl Biosci* 12: 455–462.
- Servin B, Dillmann C, Decoux G and Hospital F, 2002, MDM : a program to compute fully informative genotype frequencies in complex breeding schemes. *J. Hered.*, 93(3): 227–228.
- Young ND, and Tanksley SD, 1989. Restriction fragment length polymorphism maps and the concept of graphical genotypes. *Theor Appl Genet* 77: 95–101.