Mises en perspective

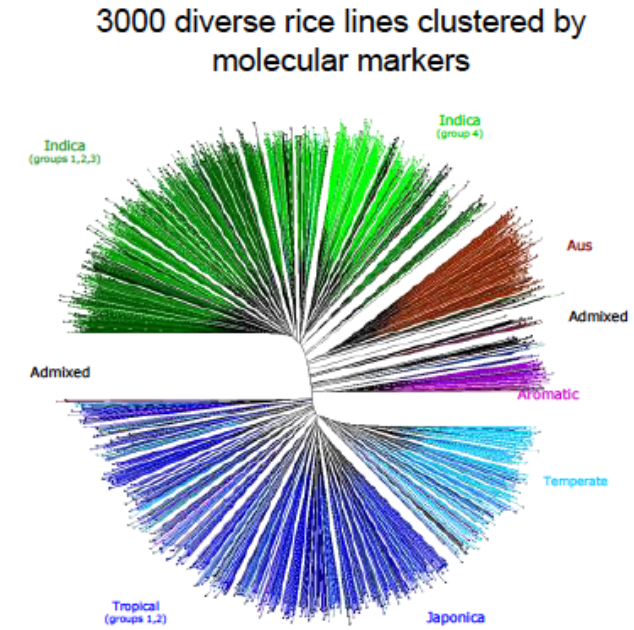
Analyse bioinformatique de séquences pour l'amélioration des plantes

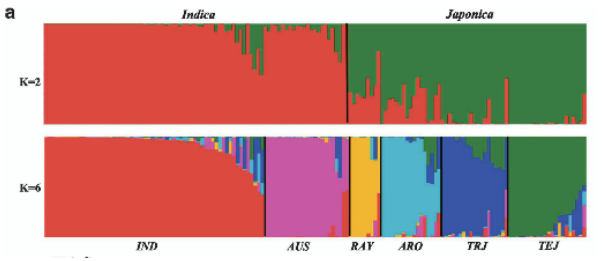
Mardi 11 février 2014

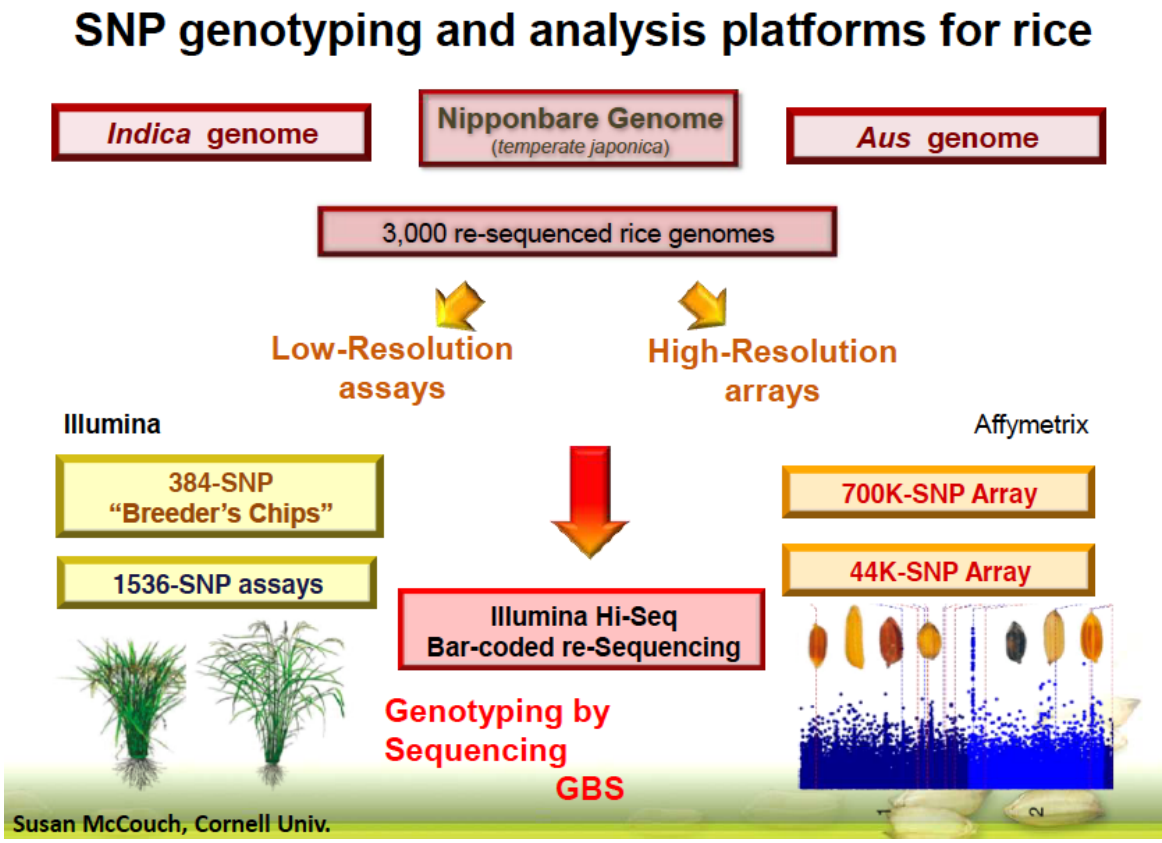
- Concepts
  - Passage de données brutes à des données informatives
    - Alignement, élimination de duplicats techniques
    - Analyse de la profondeur de lecture
    - Identification des séquences présentes dans le jeu de données
    - Présence de polymorphismes (SNP, Indels), vrais ou faux…
    - Traitement des données: LD, design d'une puce SNP haut débit (Illumina Veracode)…
  - Formats SAM (Sequence Alignment Map) / BAM (compressé), VCF (Variant Call Format)

- Compétences
  - Faire un assemblage de données NGS (MIRA)
  - Positionner ces séquences sur une référence = mapping (BWA)
  - Trier des fichiers
  - Créer un workflow
  - Détecter des SNP (GATK, SNiPlay)

Mise en perspective: Polymorphisme et sélection

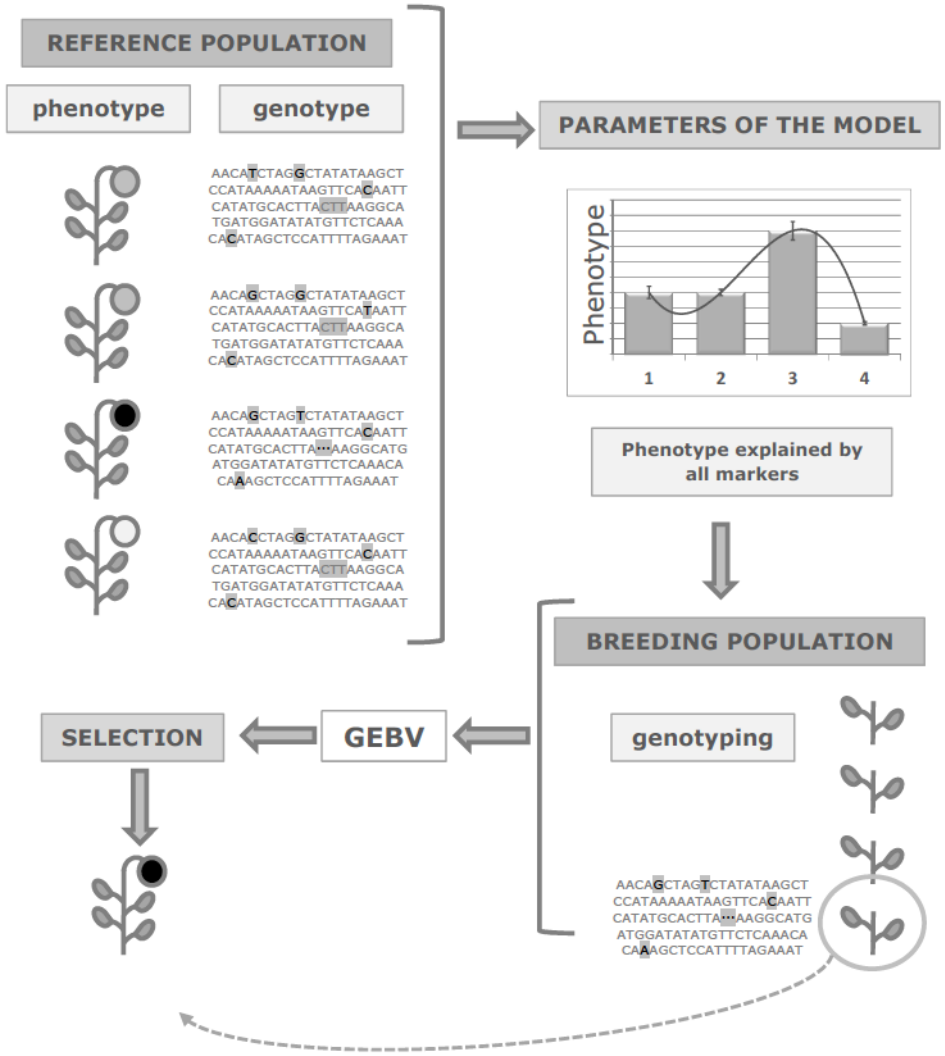






Analyses de **diversité**: Stratégie de génotypage haute définition sur 3000 génomes de riz (Genotyping by sequencing)

Mise en perspective: Polymorphisme et sélection



Sélection génomique
(Perez de Castro et al 2012)

Mise en perspective: Polymorphisme et sélection

Table 1 | **A list of available non-commercial NGS genotype-calling software**

Software	Available from	Calling method	Prerequisites	Comments	Refs
SOAP2	http://soap.genomics.org.cn/index.html	Single-sample	High-quality variant database (for example, dbSNP)	Package for NGS data analysis, which includes a single individual genotype caller (SOAPsnp)	15
realSFS	http://128.32.118.212/thorfinn/realSFS/	Single-sample	Aligned reads	Software for SNP and genotype calling using single individuals and allele frequencies. Site frequency spectrum (SFS) estimation	-
Samtools	http://samtools.sourceforge.net/	Multi-sample	Aligned reads	Package for manipulation of NGS alignments, which includes a computation of genotype likelihoods (samtools) and SNP and genotype calling (bcftools)	53
GATK	http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit	Multi-sample	Aligned reads	Package for aligned NGS data analysis, which includes a SNP and genotype caller (Unifed Genotyper), SNP filtering (Variant Filtration) and SNP quality recalibration (Variant Recalibrator)	32,33
Beagle	http://faculty.washington.edu/browning/beagle/beagle.html	Multi-sample LD	Candidate SNPs, genotype likelihoods	Software for imputation, phasing and association that includes a mode for genotype calling	42
IMPUTE2	http://mathgen.stats.ox.ac.uk/impute/impute_v2.html	Multi-sample LD	Candidate SNPs, genotype likelihoods	Software for imputation and phasing, including a mode for genotype calling. Requires fine-scale linkage map	44
QCall	ftp://ftp.sanger.ac.uk/pub/rd/QCALL	Multi-sample LD	'Feasible' genealogies at a dense set of loci, genotype likelihoods	Software for SNP and genotype calling, including a method for generating candidate SNPs without LD information (NLDA) and a method for incorporating LD information (LDA). The 'feasible' genealogies can be generated using Margarita (http://www.sanger.ac.uk/resources/software/margarita)	54
MaCH	http://genome.sph.umich.edu/wiki/Thunder	Multi-sample LD	Genotype likelihoods	Software for SNP and genotype calling, including a method (GPT_Freq) for generating candidate SNPs without LD information and a method (thunder_glf_freq) for incorporating LD information	-

A more complete list is available from http://seqanswers.com/wiki/Software/list. LD, linkage disequilibrium; NGS, next-generation sequencing.

(Nielsen et al 2011)

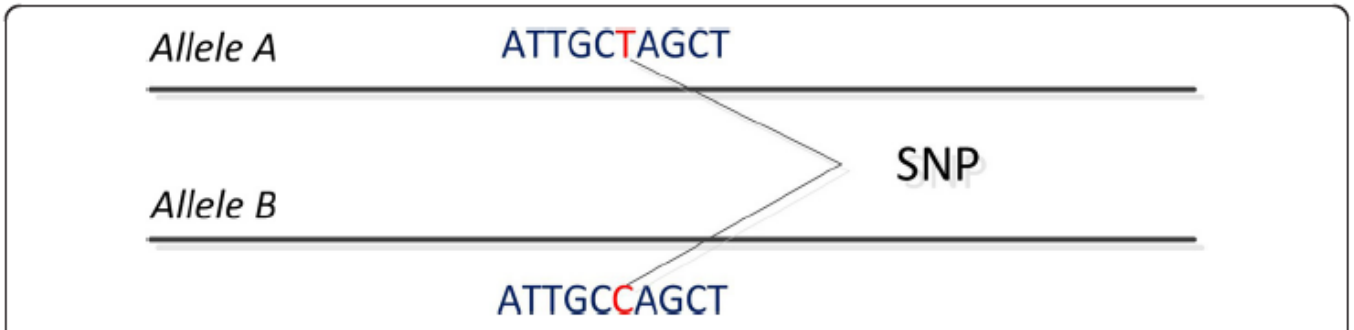


Figure 3 SNP example. A difference in a single nucleotide between two DNA fragments from different individuals. In this case we say that there are two alleles: C and T.

(Pavlopoulos et al 2013)

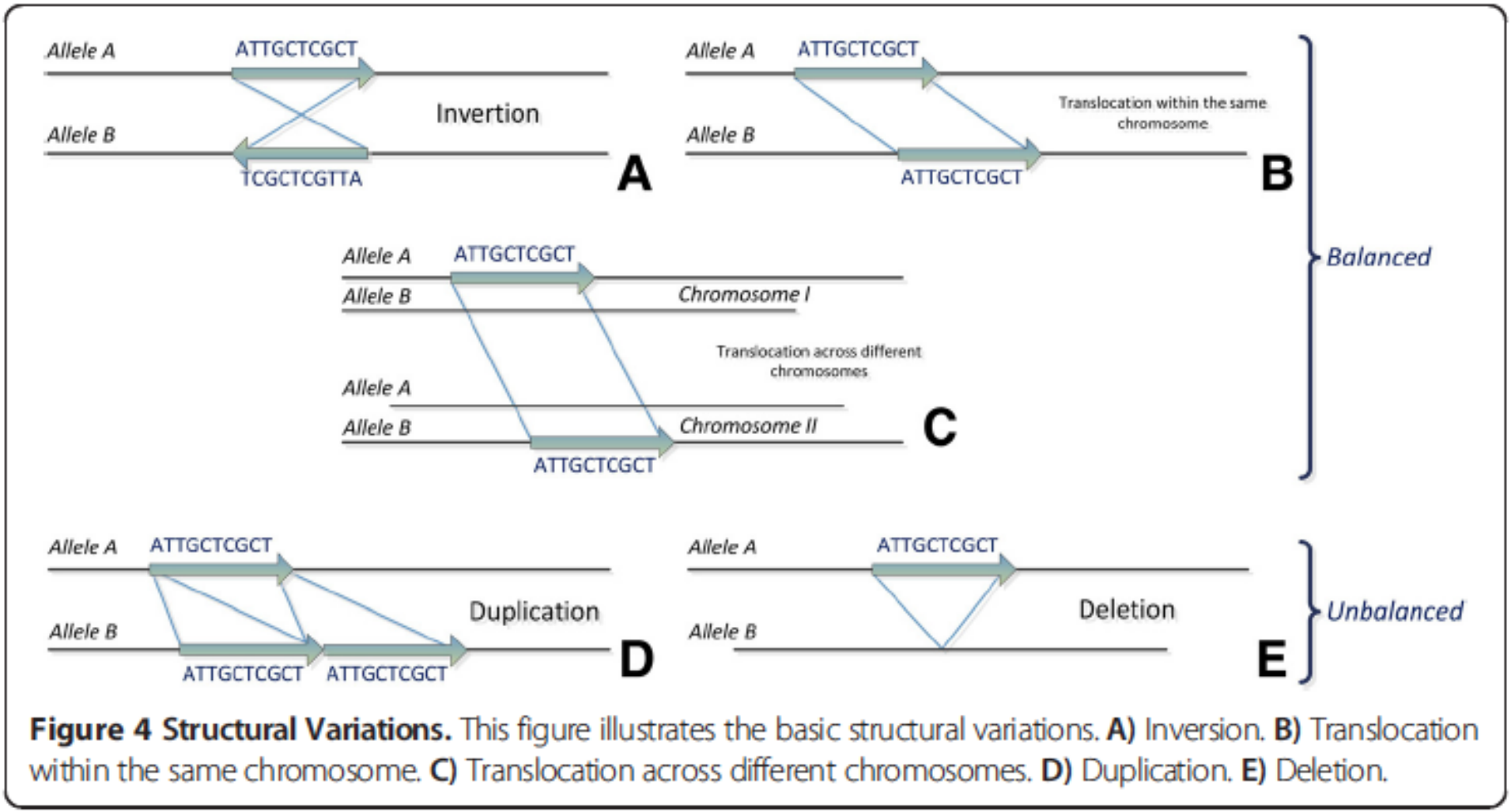


Figure 4 Structural Variations. This figure illustrates the basic structural variations. **A)** Inversion. **B)** Translocation within the same chromosome. **C)** Translocation across different chromosomes. **D)** Duplication. **E)** Deletion.

Mise en perspective: Polymorphisme et sélection

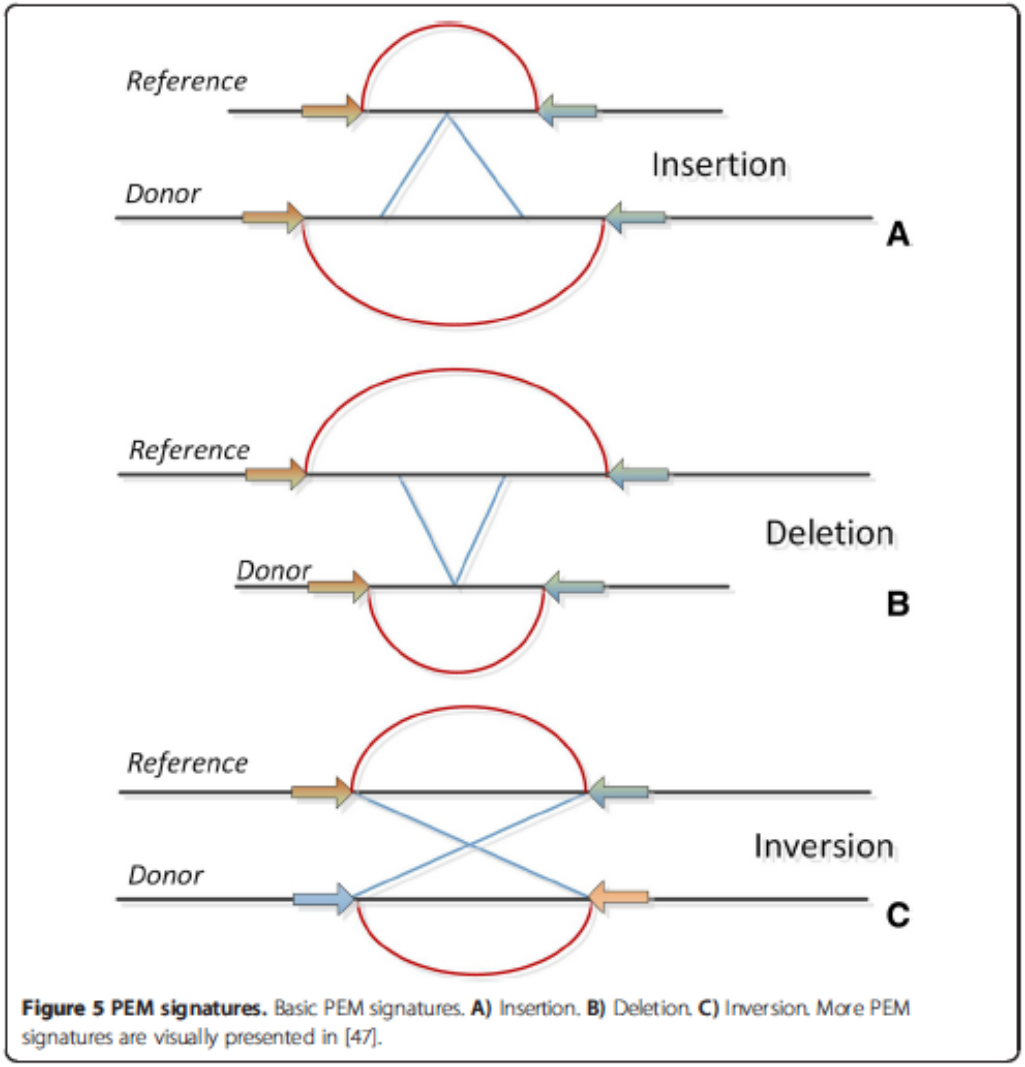


Figure 5 PEM signatures. Basic PEM signatures. **A)** Insertion. **B)** Deletion. **C)** Inversion. More PEM signatures are visually presented in [47].

(Pavlopoulos et al 2013)

(Pavlopoulos et al 2013)

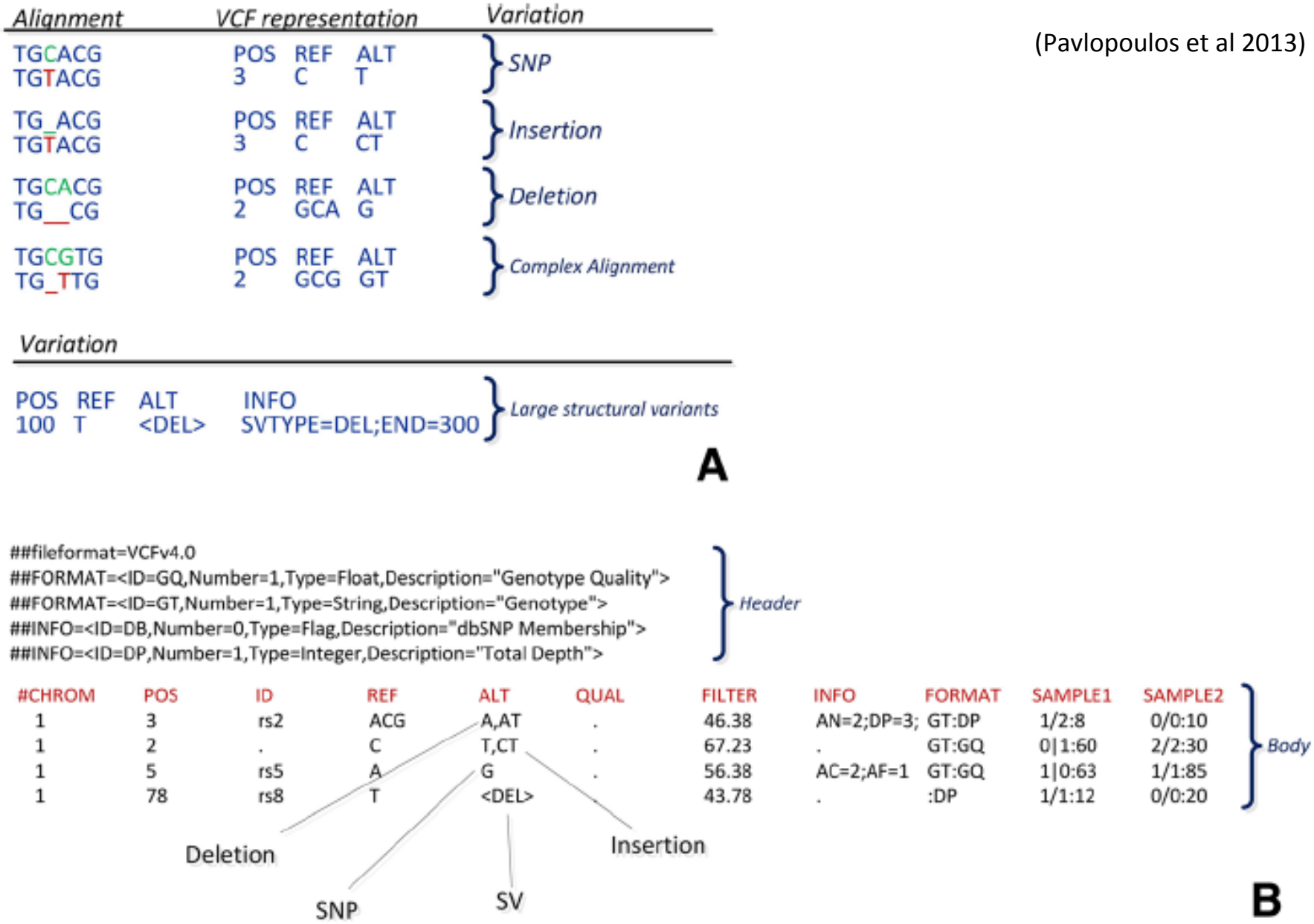


Figure 9 VCF file. This figure demonstrates an example of a CVF file. **A)** Different types of variations and polymorphisms that can be stored in CVF format. **B)** Example of a CVF format and its fields.

Applications des polymorph

Bibliographie

- G.A. Pavlopoulos et al. **Unravelling genomic variation from next generation sequencing data.** Biodata Mining (2013) 6:13
- A.M. Pérez-de-Castro et al. **Application of genomic tools in Plant breeding.** Current genomics (2012) 13, 179-195
- M.A. DePristo et al. **A framework for variation discovery and genotyping using next-generation DNA sequencing data**. Nature Genetics (2011) 43. 491-498
- R. Nielsen et al. **Genotype and SNP calling from next generation sequencing data**. Nature Genetics (2011) 12. 443-451