SNP calling and visualization of vcf files

M. Schilling

August 8, 2014

SNP calling

There are plenty of different ways to call genotypes

- Samtools
- SOAPsnp
- GATK
- GotCloud
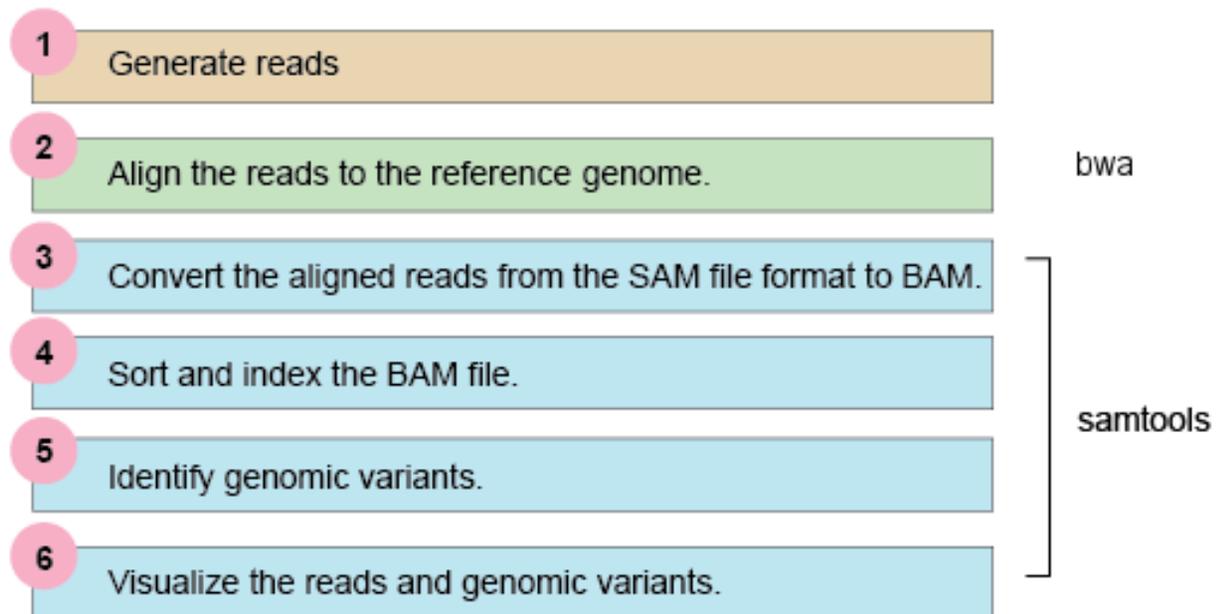- Stacks
- UNEAK
- CRoPS
- ...

SNP calling

There are plenty of different ways to call genotypes
When deciding which one to use, consider

- Which organism am I working on?
- How big is my data set?
- What am I going to do with these SNPs?
- How geeky/nerdy do I want to be?

Samtools - recap

Why so many steps?

mod. from biobits.org/samtools_primer.html

Visualization of Polymorphisms

- tview

Visualization of Polymorphisms

- tview
- Tablet
- IGV (Integrative Genomics Viewer) - kitchen sink of formats
- inPHAP (Haplotypes and SNPs)
- VARB (VARiation Browser)
- Savant Genome Browser (still maintained?)

A note about Minor Allele Frequencies

- Most nucleotides are invariant or have a rare (MAF < 0.05) segregating minor allele. (66% rare with 530k total variants for Lycaeides (Gompert et al., 2013))
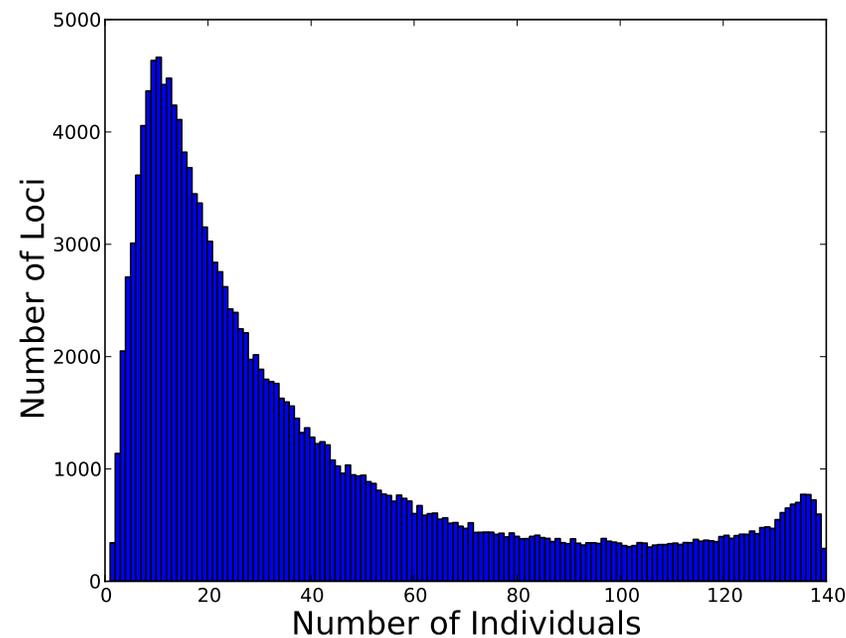
A note about Minor Allele Frequencies

- Most nucleotides are invariant or have a rare (MAF < 0.05) segregating minor allele. (66% rare with 530k total variants for Lycaeides (Gompert et al., 2013))
- These rare variants could be associated with local adaptation/differentiation
- Given this complexity, are we discarding most of the variation?

*References

Gompert, Z., Lucas, L. K., Buerkle, C. A., Forister, M. L., Fordyce, J. A., and Nice, C. C. (2013). Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. *Molecular Ecology*.