# Pedigree Reconstruction from SNPs

## Jisca Huisman



## Why SNPs?

THE UNIVERSITY of EDINBURGH

Data on thousands of Single Nucleotide Polymorphisms (SNPs) has a plethora of uses, including estimating heritability and inbreeding. They can be scored in far greater numbers than microsatellite markers, including from low-quality samples, with less human input.

### Genes vs genealogy

To infer pedigree relationships from genetic relatedness, a simple threshold can suffice to distinguish first degree relatives from unrelated pairs. However, this approach is inefficient to classify relatives in a more complex mixture, due to extensive overlap in relatedness. Likelihood methods (such as Sequoia and Colony) are more powerful, giving more correct assignments

and fewer errors.

*Emperical values* estimated from 40.000 SNPs



## Why a pedigree?

Knowing the parents of individuals provides insight into the mating system of a species. It also provides estimates of breeding success, a key fitness component.

In social species, knowing how pairs of interacting individuals are related can aid in understanding social dynamics and the role of kin recognition.

How?

0/1/2 Genotypes in Input: format (no map order needed), and a file with ID, sex and year of birth/hatching.

Parentage: 'Sieve' through all possible pairs, first using #SNPs at which opposing homozythen using likelihood gotes, (LLR) parent-offspring ratio versus unrelated. Assign if parent-offspring likelihood exceeds likelihood of being a different kind of 1<sup>st</sup>, 2<sup>nd</sup> or 3<sup>rd</sup> degree relative.

#### using GCTA. Unrelated Showing equal proportions of the categories. 0.6 -0.1 0.1 0.2 0.4 0.5 0.3 $\mathbf{0}$ 0.7 Genomic relatedness

Why sequoia?

Datasets from wild populations are often characterised by incomplete sampling and incomplete knowledge of genders and ages. Sequoia accomodates such real-world datasets, as well as polyga-mous mating systems with overlapping generations and various kinds of inbreeding.

Genotyped male • Genotyped female ○ Non-genotyped Only paternal links shown

Parent assignment ... of genotyped parents to genotyped individuals Sibship clustering Assign 'dummy' parents to groups of half-siblings Grandparent

Connect sibships to the remainder of the pedigree

#### Avuncular

Connect non-genotyped parents

Full pedigree: Sieve using LLR focal relation-ship / unrelated, assign if focal relationship gives higher likelihood than other 1<sup>st</sup>, 2<sup>nd</sup> or 3<sup>rd</sup> degree relationships, calculated over the pair and their currently-assigned close relatives



#### Performance

Sequoia generally has high assignment rates (>99%), low error rates (<1%), and fast runtime (often <10 min). Colony has lower error rates, especially when using  $\leq 200$  SNPs, but its use when generations Method (200 SNPs) overlap is not straight Sequoia – parentage Sequoia - full forward.



generation

1.0 -

0.8

0.6

0.4

0.2

0.0

Assignment rate



#### Example in Rum red deer



This long term study population formed the inspiration to write Sequoia. Shown above is the ancestry of `Cindy', the collared calf pictured in the photo below with both her parents. Her non-genotyped father was presumed to be an immigrant, but Sequoia revealed that he originated from within the study population. The resulting ancestry has many loops in the older, less well sampled period. ISLE OF RUM

#### Acknowledgements

I would like to thank J Pemberton for her support, B Hill, J Hadfield and 3 anonymous reviewers for comments and discussion, G Gorjanc for help building the R package, and P Bijma for handing me a Fortran book many years ago.



#### Prop. genotyped parents

Pedigrees contain 1157, 960 and 1998 genotyped offspring respectively, and 200 independent SNPs were simulated 10x. Red deer pedigree contains some close inbreeding and few unknown *birth years. Default analysis parameter used.* 

