

### Sub-Saharan genetic diversity and San and Khoe

Mattias Jakobsson mattias.jakobsson@ebc.uu.se

Uppsala University

H3Africa, Johannesburg, Oct 6, 2013





#### UPPSALA UNIVERSITET

### Collaborative effort







 mtDNA and Y-chromosome data indicate deepest divergences among humans.

 Sequencing of one Khoe-San individual revealed much unknown genetic variation (Schuster *et al., Nature* 2010)

• Rich ethnic diversity, including click-languages.

### Current distribution of the Khoe-San groups





### Current distribution of the Khoe-San groups





## A genome-wide study of the Khoe and San



- 7 Khoe and San groups, 4 additional SA populations
- 220 unrelated individuals
- 2,293,320 high-quality autosomal and X-chrom SNPs
- In total, 504, 530, 400 genotypes
- Data combined with HGDP, Henn et al, HapMap:
  - 269,317 autosomal SNPs
  - 1,745 unrelated individuals



Schlebusch et al., Science (2012)

## Diversity, LD, and runs of homozygosity





#### Schlebusch et al., Science (2012)

Mattias Jakobsson (UU)

Genomics of San and Khoe

# Jakobsson [ab

## Structure among sub-Saharan populations

- Procrustes correlation: 0.585
- Procrustes correlation excluding Bantu-speaking groups: 0.715



### Population divergence





Mattias Jakobsson (UU)

## Southern African sample (2.3M SNPs)













### Genome local patterns (extreme iHS)





Schlebusch et al., Science (2012)

Mattias Jakobsson (UU)

### Genome local patterns (extreme *iHS*)





Schlebusch et al., Science (2012)

Mattias Jakobsson (UU)

Genomics of San and Khoe



#### Incorporate (unbiased) genomic variation from all southern African people in biomedical infrastructure

- Discover variants by sequencing
- Study design: high/low-coverage? Exome? Pooled samples?
- Great diversity (multiple lines of evidence) can help decipher genetic basis of human traits



- Incorporate (unbiased) genomic variation from all southern African people in biomedical infrastructure
- Discover variants by sequencing
- Study design: high/low-coverage? Exome? Pooled samples?
- Great diversity (multiple lines of evidence) can help decipher genetic basis of human traits



- Incorporate (unbiased) genomic variation from all southern African people in biomedical infrastructure
- Discover variants by sequencing

#### Study design: high/low-coverage? Exome? Pooled samples?

 Great diversity (multiple lines of evidence) can help decipher genetic basis of human traits



- Incorporate (unbiased) genomic variation from all southern African people in biomedical infrastructure
- Discover variants by sequencing
- Study design: high/low-coverage? Exome? Pooled samples?
- Great diversity (multiple lines of evidence) can help decipher genetic basis of human traits

### The Jakobsson Lab



#### www.ebc.uu.se/Research/IEG/evbiol/research/Jakobsson/



#### Collaborators

- Himla Soodyall (U Witwatersrand)
- Andy Singleton (NIH)
- Michael Blum (U Grenoble)
- Flora Jay (UC Berkley)

### The Jakobsson Lab



#### www.ebc.uu.se/Research/IEG/evbiol/research/Jakobsson/



#### Funding

- ERC Starting Grant
- Swedish Research Council (VR)
- Swedish Research Council FORMAS
- Swedish Found. for International Cooperation in Research (STINT)

