MLE for Individual Ancestries Population Covariances and Selection

PhD Student: Jade Y. Cheng
Institution: Bioinformatics Research Centre
Department of Computer Science · Aarhus University

Supervisor: Rasmus Nielsen
Institution: Center for Theoretical Evolutionary Genomics
University of California Berkeley

Admixture with Quadratic Programming

$$\ln\left[P_1\left(Q,\ F\right)\right] = \sum_i^I \sum_j^J \left\{ g_{ij} \cdot \ln\left[\sum_k^K q_{ik} \cdot f_{kj}\right] + (2 - g_{ij}) \cdot \ln\left[\sum_k^K q_{ik} \cdot (1 - f_{kj})\right] \right\}.$$

Likelihood Model

$$\max_{\Delta_{Q_i}} \left\{ \frac{1}{2} \Delta_{Q_i}^T H_{Q_i} \Delta_{Q_i} + D_{Q_i}^T \Delta_{Q_i} \right\}$$

$$\text{s.t.} \quad A\Delta_{Q_i} \leq a$$

$$B\Delta_{Q_i} = b$$

$$\max_{\Delta_{F_j}} \left\{ \frac{1}{2} \Delta_{F_j}^T H_{F_j} \Delta_{F_j} + D_{F_j}^T \Delta_{F_j} \right\}$$

$$\text{s.t.} \quad A\Delta_{F_j} \leq a$$

QPAS

Find a feasible starting point
Initialize the corresponding active set
Repeat
  Solve the equality problem defined by the active set
  Compute the Lagrange multipliers of the active set
  If the solved approximation is within the feasible region
    If all Lagrange multipliers are negative
      Return the solved approximation
    Else
      Remove the constraint with the largest Lagrange multiplier
    End If
  Else
    Take the shortest step back into the feasible region
    Insert the corresponding constraint into the active set
  End If
End Repeat

Quadratic Programming

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ \mathcal{L} \end{bmatrix} = \begin{bmatrix} -D \\ b \end{bmatrix}$$

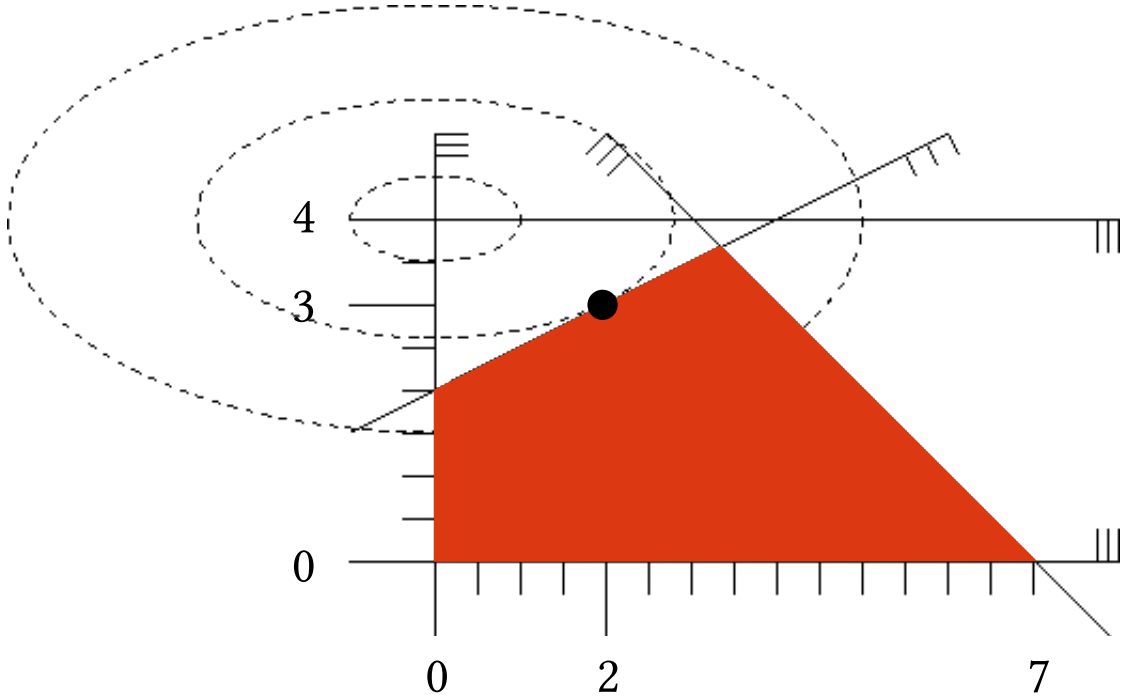
$$\max_{\Delta} \left\{ \frac{1}{2} x^T H x + D x \right\}$$

$$\text{s.t.} \quad Ax = b$$

Karush-Kuhn-Tucker
KKT EQP solover

$$\max_{\Delta} \left\{ x^2 + 4(y-4)^2 \right\}$$

$$\begin{aligned} \text{s.t.} \qquad x + y &\leq 7 \\ -x + 2y &\leq 4 \\ -x &\leq 0 \\ -y &\leq 0 \\ y &\leq 4 \end{aligned}$$



Block Relaxation

$$H_Q = \begin{bmatrix} * & * & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 & 0 \\ 0 & 0 & * & * & 0 & 0 \\ 0 & 0 & * & * & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & * & * \end{bmatrix} \qquad H_F = \begin{bmatrix} * & * & 0 & 0 & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & * & * & 0 & 0 & 0 & 0 \\ 0 & 0 & * & * & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & 0 & 0 \\ 0 & 0 & 0 & 0 & * & * & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & * & * \end{bmatrix}$$

$$\Theta\left(I^2K^2 \cdot (I + 2IK) + J^2K^2 \cdot 2JK\right)$$

$$D_Q = \begin{bmatrix} * & * & * & * & * & * \end{bmatrix} \qquad D_F = \begin{bmatrix} * & * & * & * & * & * & * & * \end{bmatrix}$$

$$H_{Q_i} = \begin{bmatrix} * & * \\ * & * \end{bmatrix} \qquad \forall i \in \{0,\ 1,\ ,\cdots,\ I-1\}$$

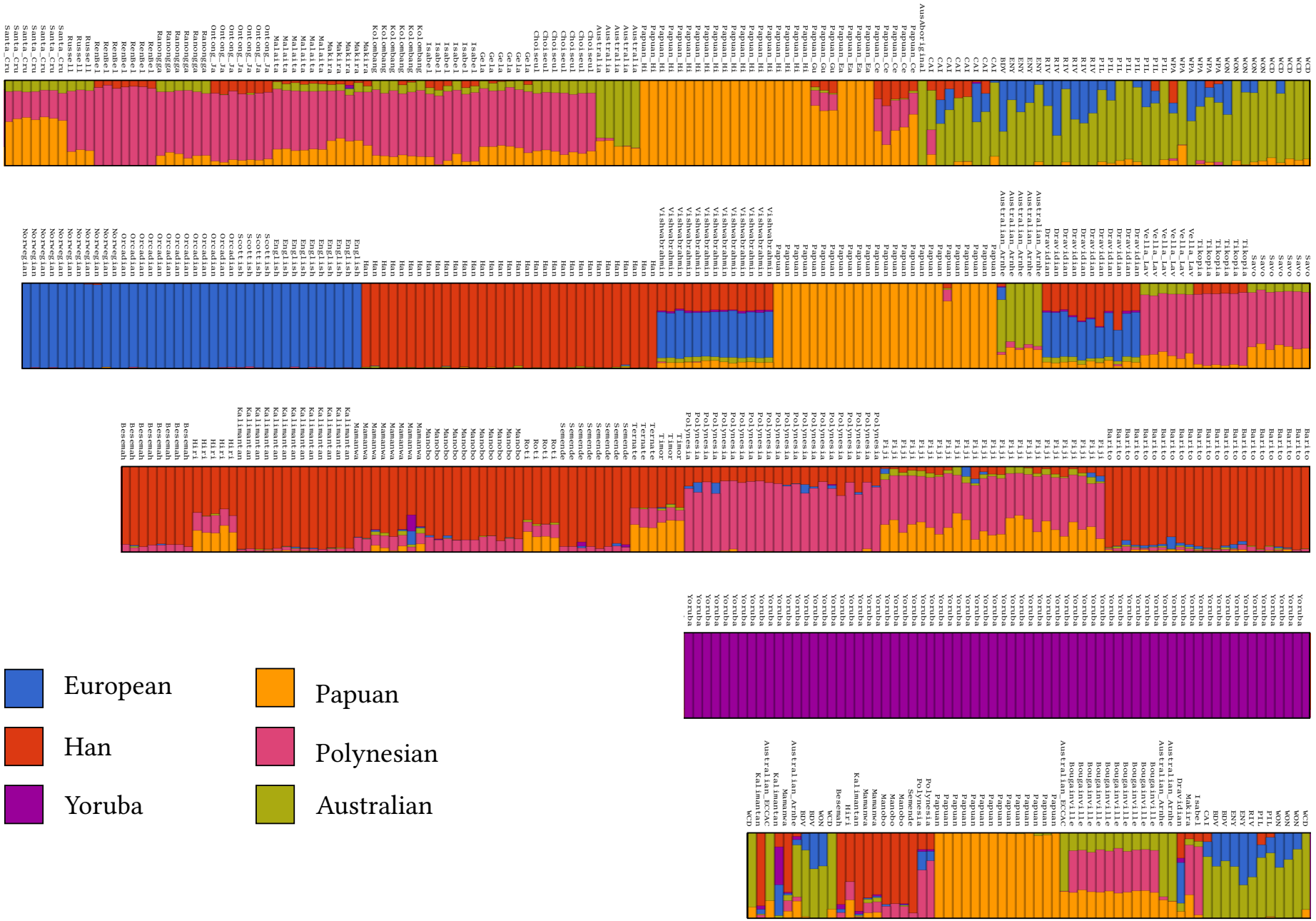
$$D_{Q_i} = \begin{bmatrix} * & * \end{bmatrix}$$

$$\Theta\left(IK^2 \cdot (1 + 2K) + JK^2 \cdot 2K\right)$$

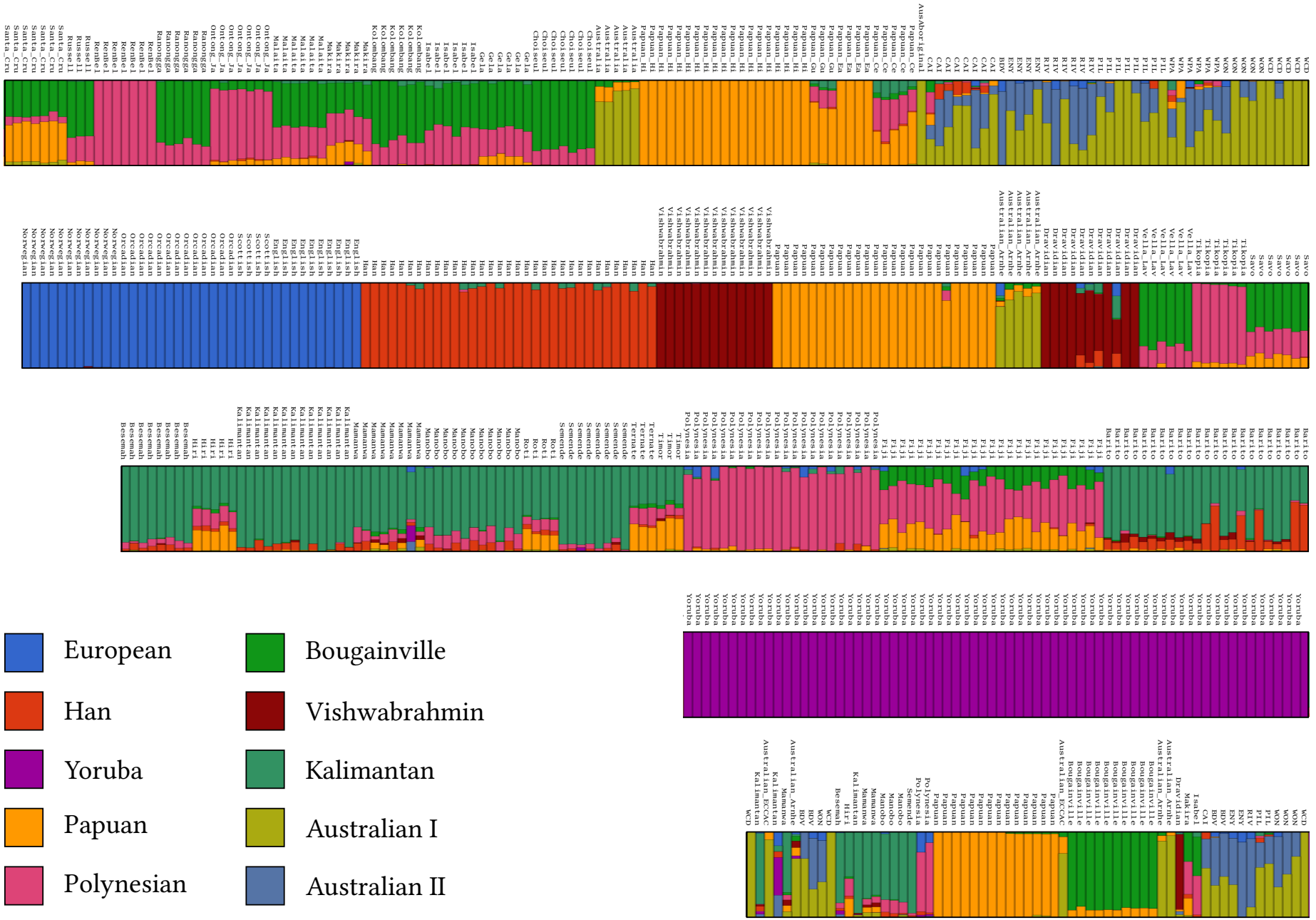
$$H_{F_j} = \begin{bmatrix} * & * \\ * & * \end{bmatrix} \qquad \forall j \in \{0,\ 1,\ ,\cdots,\ J-1\}$$

$$D_{F_j} = \begin{bmatrix} * & * \end{bmatrix}$$

Admixture Analysis



Admixture Analysis



Jade Cheng · BiRC · Aarhus University · September 10, 2015

Admixture Comparison with Other Software

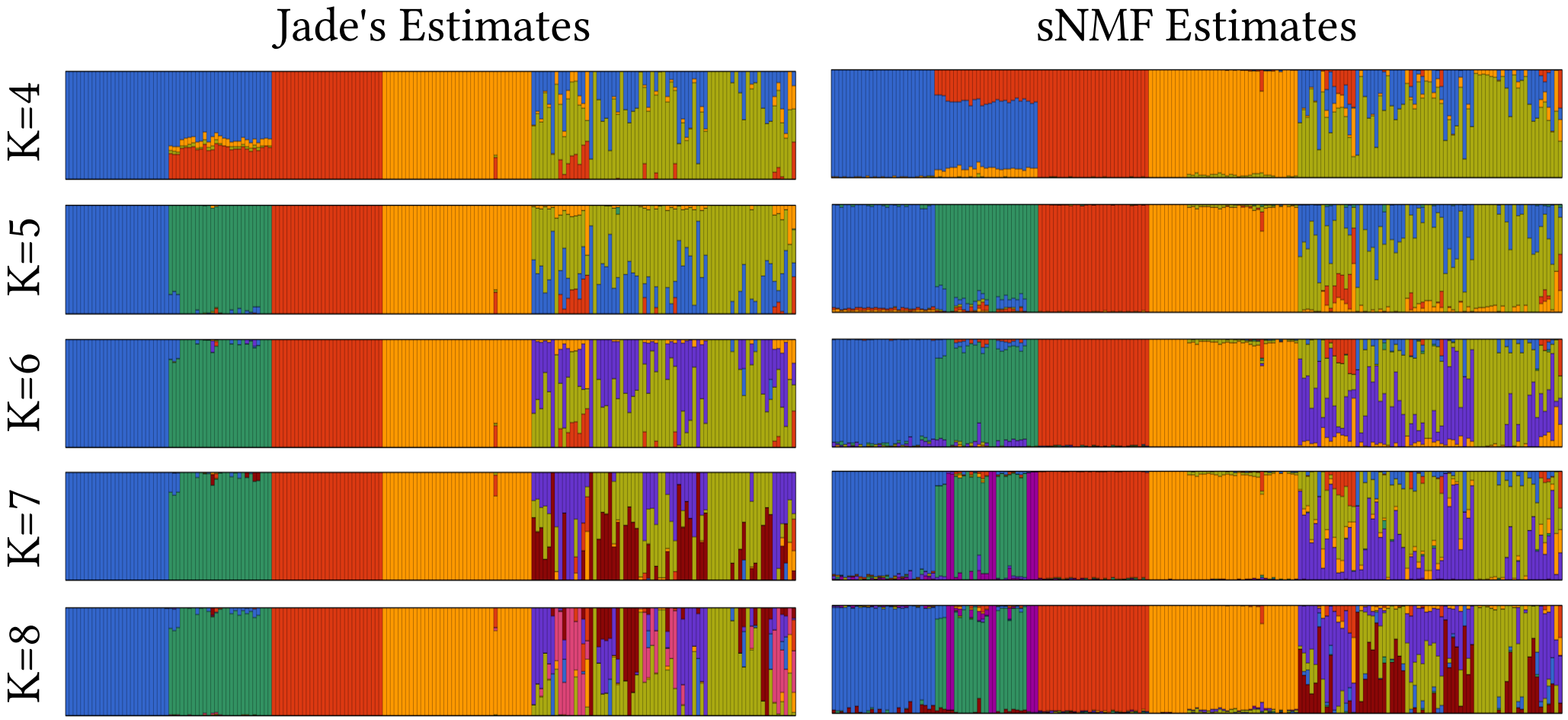
Dataset #1: 104,453 markers and 571 samples.

K	log likelihood Jade's estimates	log likelihood ADMIXTURE estimates	difference
4	-53554440.2845700	-53554561.351797	121.067227050700
5	-53234096.2234545	-53234223.237922	127.014467492700
6	-53100997.2949234	-53123544.019126	22546.7242026329
7	-52978925.4104690	-53011655.087574	32729.6771049723

Dataset #2: 566,359 markers and 191 samples.

K	log likelihood Jade's estimates	log likelihood sNMF estimates	difference
4	-66320990.634187	-66556930.5221196	235939.887932852
5	-65489775.091462	-65681252.2277625	191477.136300825
6	-65117441.449915	-65300580.5042625	183139.054347418
7	-64792022.992634	-64955168.6613403	163145.668706432
8	-64493170.693910	-64675411.6970119	182241.003101580

Admixture Comparison with Other Software



Constraint Admixture Inference

```
# 'I' Individuals
152

# 'K' Populations
5

# Population Assignments per Individual
3 3 3 3 3
3 3 3 3 3 3 3 3
4 4 4 4 4 4 4
4 4 4 4 4
4 4 4 4 4 4 4 4 4 4
3 3 3 3 3 3 3 3
4 4 4 4 4 4 4 4
4 4 4 4 4 4 4 4
3 3 3
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
1 1 1 1 1 1 1 1 1 1
0 0 0 0 0
2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0
0 0 0 0 0
```

```
# Population 0 af
10 1
0.400000
0.000000
0.000000
0.000000
0.000000
1.000000
1.000000
1.000000
1.000000
1.000000

# Population 1 eu
10 1
0.000000
0.400000
0.000000
0.000000
0.000000
0.000000
1.000000
1.000000
1.000000
1.000000
```

```
# Population 2 han
10 1
0.000000
0.000000
0.400000
0.000000
0.000000
1.000000
1.000000
1.000000
1.000000
1.000000

# Population 3 au1
10 1
0.000000
0.000000
0.000000
0.400000
0.000000
1.000000
1.000000
1.000000
1.000000
0.000001
```

```
# Population 4 au2
10 1
0.000000
0.000000
0.000000
0.000000
0.400000
1.000000
1.000000
1.000000
0.000001
1.000000
```

Constraint Admixture Configuration File

Constraint Admixture Inference