The effect of population size change on the SFS

In[51]:=
```
Date[]
```

Out[51]=
```
{2020, 12, 21, 22, 47, 55.079822}
```

Piecewise code (by Maxim Rytin):

Download and execute the initialization cells of the Mathematica notebook “piecewise.nb” from https://library.wolfram.com/infocenter/MathSource/5117/

Johri 2020 preprint - analytical expectation w/o selection

To replicate the results of Johri et al., (after running the piecewise code referred to above) run all initialization cells of the notebook. On an i5-7500T processor with up to 3.30 GHz, and running Mathematica 12.1, this took somewhat over an hour (or 1.5 hours on a 2017 MacBook 12-inch with i5 CPU). The results presented in the manuscript are computed in the last section “Including the effects of BGS - Different values of B”.

The demographic scenarios:

Supp Table 8: Parameters underlying the human-like demographic models considered.

	Demographic models	**Ancestral population size**	**Current population size**	**Time of change in generations**
1	Equilibrium	10,000	10,000	NA
2	Exponential growth	1000	30,000	850
3	Instantaneous decline	12,300	2,100	4,750

Expressions from Polanski and Kimmel. 2003

The calculations below follow the approach of Polanski, A., and Kimmel, M. (2003). New Explicit Expressions for Relative Frequencies of Single-Nucleotide Polymorphisms With Application to Statistical Inference on Population Growth.” *Genetics* 165:427–436. The equation numbers used by Polanski and Kimmel (2003) are indicated here and in the exponential growth section.

Equation 6 (coefficient used in equations 9 and 10):

In[52]:=
```
Akjn[k_,j_,n_]:= Product[Binomial[l,2],{l,Select[Range[k,n],#≠j&]}] / Product[Binomial[l,2]-Binomial[j,2],{l,Select[Range[k,n],#≠j&]}]
```

Equation 9 (coefficient used in equation 8):

In[53]:=
```
vnj[n_, j_] := Sum[j (j - 1) Akjn[k, j, n] / (k - 1), {k, 2, j}]
```

Equation 10 (coefficient used in equation 8):

In[54]:=
```
wnbj[n_, b_, j_] := Sum[j (j - 1) Binomial[n - k, b - 1] ((n - b - 1) ! (b - 1) !) / (n - 1) ! Akjn[k, j, n], {k, 2, j}]
```

Exponential growth model

A function to generate an expression for the effective population size depending on time (t in generations) under the expansion scenario (corresponds to equation 5):

In[55]:=
```
NeGenExp[anf_, end_, tchange_] := Piecewise[{{anf (end/anf)^((tchange-t)/tchange), t < tchange}, {anf, t ≥ tchange}}]
```

The resulting expression/the demographic scenario.

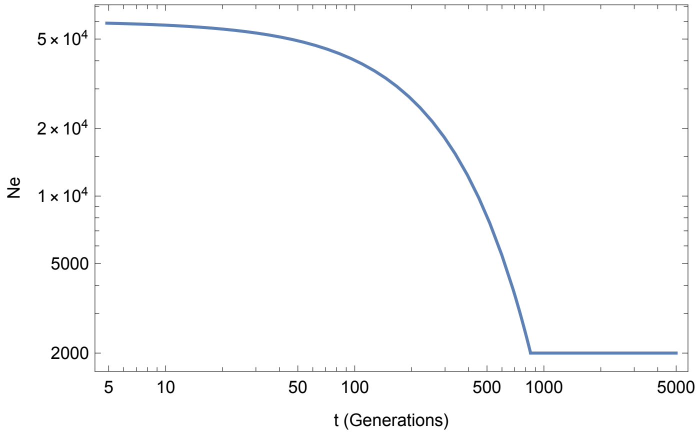
In[56]:=
```
NGE = NeGenExp[2 × 1000, 2 × 30 000, 850]
```

Out[56]=

$$\begin{cases} 2^{4+\frac{850-t}{850}} \times 3^{\frac{850-t}{850}} \times 5^{3+\frac{850-t}{850}} & t < 850 \\ 2000 & t \geq 850 \\ 0 & \text{True} \end{cases}$$

In[57]:=
```
LogLogPlot[NGE, {t, 0, 5000}, AxesOrigin → {0, 0}, Frame → True, FrameLabel → {"t (Generations)", "Ne"}]
```

Out[57]=



Equation 4 (distribution of time to coalescence, given the demographic scenario NGE - exponential increase):

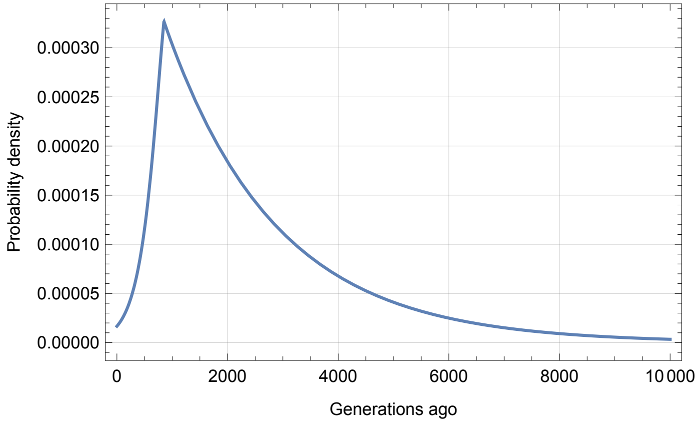
In[58]:=

$$\texttt{qjtGenExp[j\_,B\_]:=}\frac{\texttt{Binomial[j,2]}}{\texttt{B NGE}}\texttt{Exp}\left[\texttt{-PiecewiseIntegrate}\left[\frac{\texttt{Binomial[j,2]}}{\texttt{B NGE/.t}\rightarrow\sigma},\{\sigma,0,t\}\right]\right]$$

Coalescence time for a pair of lineages under expansion model without BGS:

In[59]:=
```
Plot[qjtGenExp[2, 1], {t, 0, 10 000}, Frame -> True,
 FrameLabel → {"Generations ago", "Probability density"}, GridLines → Automatic]
```

Out[59]=



Equation 3 (the expectation of equation 4):

In[60]:=
```
ejGenExp[j_,B_]:=PiecewiseIntegrate[t qjtGenExp[j,B],{t,0,∞}]
```

Expected coalescence time for a pair of lineages under expansion model without BGS.

In[61]:=
```
ejGenExp[2, 1] // N
```

Out[61]= 2596.63

Equation 8 (an expression for the probability to see *b* derived alleles in a sample of *n* with the effective population size scaled by *B*):

In[62]:=
```
qnbGenExpL[n_, b_, B_] := Sum[ejGenExp[j, B] × wnbj[n, b, j], {j, 2, n}] / Sum[ejGenExp[j, B] × vnj[n, j], {j, 2, n}]
```

As a test, derive an expression where B is set to 1 (no BGS). This does not have to be executed, and will be replicated in the last section (Including the effects of BGS - different values of B).

Instantaneous decline model

This is analogous to the section above, but adjusted for the decline model.

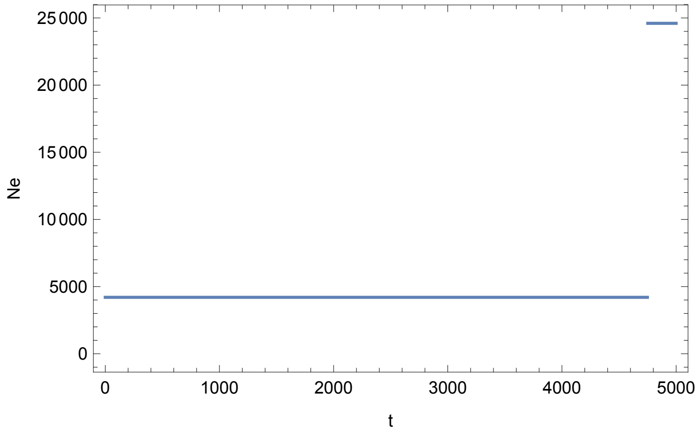
In[63]:=
```
NeGenStep[anf_, end_, tchange_] := Piecewise[{{anf, t < tchange}, {end, t ≥ tchange}}]
```

In[64]:=
```
NGS = NeGenStep[2 × 2100, 2 × 12 300, 4750]
```

Out[64]=
```
4200    t < 4750
24 600  t ≥ 4750
0       True
```

In[65]:=
```
Plot[NGS, {t, 0, 5000}, AxesOrigin → {0, 0}, Frame → True, FrameLabel → {"t", "Ne"}]
```

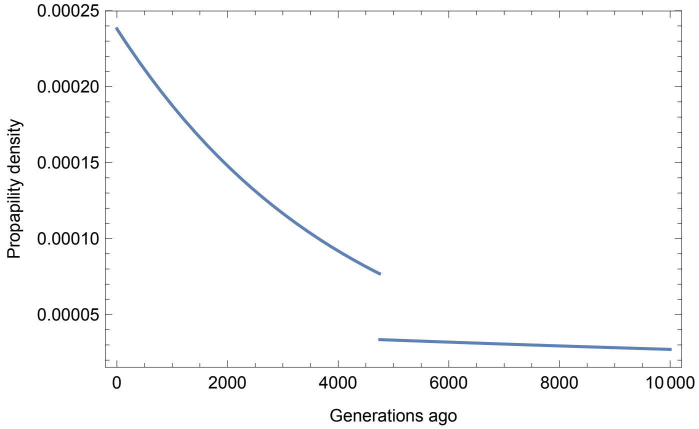
Out[65]=



In[66]:=
```
qjtGenStep[j_,B_]:= Binomial[j,2]/(NGS B) Exp[-PiecewiseIntegrate[Binomial[j,2]/(B NGS/.t→σ),{σ,0,t}]]
```

Coalescence probability for a pair of lineages under the step change model:

In[67]:=
```
Plot[qjtGenStep[2, 1], {t, 0, 10 000}, Frame → True, FrameLabel → {"Generations ago", "Propapility density"}]
```

Out[67]=


In[68]:=
```
ejGenStep[j_,B_]:=PiecewiseIntegrate[t qjtGenStep[j,B],{t,0,∞}]
```

Expectation of the coalescence probability for a pair of lineages under the step change model:

In[∘]:=
```
ejGenStep[2, 1] // N
```

Out[∘]=
```
10 783.6
```

In[69]:=
```
qnbGenStepL[n_, b_, B_] := Sum[ejGenStep[j, B] × wnbj[n, b, j], {j, 2, n}] / Sum[ejGenStep[j, B] × vnj[n, j], {j, 2, n}]
```

In[∘]:=
```
AbsoluteTiming[qnbGenStep20L = Simplify[qnbGenStepL[20, b, 1], TimeConstraint → 1];]
```

Out[∘]=
```
{19.0018, Null}
```

In[∘]:=
```
ByteCount[qnbGenStep20L]
```

Out[∘]=
```
71 888
```

In[•]:=
```
Table[qnbGenStep20L, {b, 1, 19}] // N
```
Out[•]=
```
{0.161012, 0.0964657, 0.0746265, 0.0634678, 0.0565841, 0.0518402, 0.0483211, 0.0455693, 0.0433305,
 0.0414522, 0.0398374, 0.0384213, 0.037159, 0.0360187, 0.0349767, 0.0340155, 0.0331216, 0.0322845, 0.0314959}
```

Constant size model

In[70]:=
$$qjtGen10k[j\_,B\_] := \frac{Binomial[j,2]}{2 \times 10000B} Exp\left[-PiecewiseIntegrate\left[\frac{Binomial[j,2]}{2 \times 10000B},\{\sigma,0,t\}\right]\right]$$

In[71]:=
```
ejGen10k[j_,B_]:=PiecewiseIntegrate[t qjtGen10k[j,B],{t,0,∞}]
```

In[72]:=
$$qnb10kL[n\_,b\_,B\_] := \frac{Sum[ejGen10k[j,B] \times wnbj[n,b,j],\{j,2,n\}]}{Sum[ejGen10k[j,B] \times vnj[n,j],\{j,2,n\}]}$$

In[•]:=
```
AbsoluteTiming[qnb10kL20 = Simplify[qnb10kL[20, b], TimeConstraint → 1];]
```
Out[•]=
```
{0.000088, Null}
```

In[•]:=
```
AbsoluteTiming[qnb10kL20 = Simplify[qnb10kL[20, b, 1], TimeConstraint → 1];]
```
Out[•]=
```
{3.59072, Null}
```

In[•]:=
```
Table[qnb10kL20, {b, 1, 19}] // N
```
Out[•]=
```
{0.28187, 0.140935, 0.0939565, 0.0704674, 0.0563739, 0.0469783, 0.0402671, 0.0352337, 0.0313188,
 0.028187, 0.0256245, 0.0234891, 0.0216823, 0.0201335, 0.0187913, 0.0176169, 0.0165806, 0.0156594, 0.0148352}
```

Including the effects of BGS - Different values of B

Post-burn-in values of B

The genetic diversity at the end of the simulations is still more or less in-line with the levels of BGS before the demographic change started. Using values of B obtained from the simulations (after burn-in, but before population size change), allows good predictions of the final levels of nucleotide diversity and

B.

In[73]:=
```
bPostBurnIn=Rationalize[{{1.000,0.716,0.602,0.253,0.540,0.388,0.568},
 {1.000,0.710,0.602,0.252,0.538,0.386,0.568},
 {1.000,0.856,0.827,0.790,0.850,0.871,0.920}}];
```

In[•]:=
```
TableForm[bPostBurnIn,
 TableHeadings → {{"Constant", "Decline", "Expansion"}, {"Neutral", "DFE1", "DFE2", "DFE3", "DFE4", "DFE5", "DFE6"}}]
```

Out[•]//TableForm=

	Neutral	DFE1	DFE2	DFE3	DFE4	DFE5	DFE6
Constant	1	$\frac{179}{250}$	$\frac{301}{500}$	$\frac{253}{1000}$	$\frac{27}{50}$	$\frac{97}{250}$	$\frac{71}{125}$
Decline	1	$\frac{71}{100}$	$\frac{301}{500}$	$\frac{63}{250}$	$\frac{269}{500}$	$\frac{193}{500}$	$\frac{71}{125}$
Expansion	1	$\frac{107}{125}$	$\frac{827}{1000}$	$\frac{79}{100}$	$\frac{17}{20}$	$\frac{871}{1000}$	$\frac{23}{25}$

Derive expressions for the site-frequency spectra (depending on B) for each demographic scenario:

In[74]:=
```
AbsoluteTiming[termsConsPBI = Table[qnb10kL[20, b, bPostBurnIn[[1, k]]], {k, 1, 7}];]
```

Out[74]= {31.1513, Null}

In[75]:=
```
AbsoluteTiming[termsStepPBI = Table[qnbGenStepL[20, b, bPostBurnIn[[2, k]]], {k, 1, 7}];]
```

••• PolynomialGCD: Exponent is out of bounds for function PolynomialGCD.

••• PolynomialGCD: Exponent is out of bounds for function PolynomialGCD.

••• PolynomialGCD: Exponent is out of bounds for function PolynomialGCD.

••• General: Further output of PolynomialGCD::lrgexp will be suppressed during this calculation.

Out[75]= {495.525, Null}

In[76]:=
```
AbsoluteTiming[termsExpPBI = Table[qnbGenExpL[20, b, bPostBurnIn[[3, k]]], {k, 1, 7}];]
```

Out[76]= {4645.8, Null}

In[•]:=
```
ByteCount[termsConsPBI]
ByteCount[termsStepPBI]
ByteCount[termsExpPBI]
```

Out[•]= 810 096

Out[•]= 899 624

Out[•]= 1 258 968

Get site-frequency spectra for a sample of size 20 (constant, step decline, exponential expansion). Rows: neutral, DFE1, DFE2, DFE3, DFE4, DFE5, DFE6. Cols: Frequency classes 1-19

Constant population size

In[77]:=
```
TableForm[Table[termsConsPBI[[k]], {k, 1, 7}, {b, 1, 19}] // N,
 TableHeadings → {{"Neutral", "DFE1", "DFE2", "DFE3", "DFE4", "DFE5", "DFE6"}, Range[1, 19]}]
```

Out[77]//TableForm=

	1	2	3	4	5	6	7	8	9	10
Neutral	0.28187	0.140935	0.0939565	0.0704674	0.0563739	0.0469783	0.0402671	0.0352337	0.0313188	0.0
DFE1	0.28187	0.140935	0.0939565	0.0704674	0.0563739	0.0469783	0.0402671	0.0352337	0.0313188	0.0
DFE2	0.28187	0.140935	0.0939565	0.0704674	0.0563739	0.0469783	0.0402671	0.0352337	0.0313188	0.0
DFE3	0.28187	0.140935	0.0939565	0.0704674	0.0563739	0.0469783	0.0402671	0.0352337	0.0313188	0.0
DFE4	0.28187	0.140935	0.0939565	0.0704674	0.0563739	0.0469783	0.0402671	0.0352337	0.0313188	0.0
DFE5	0.28187	0.140935	0.0939565	0.0704674	0.0563739	0.0469783	0.0402671	0.0352337	0.0313188	0.0
DFE6	0.28187	0.140935	0.0939565	0.0704674	0.0563739	0.0469783	0.0402671	0.0352337	0.0313188	0.0

Step decline model

In[78]:=

```
TableForm[Table[termsStepPBI[[k]], {k, 1, 7}, {b, 1, 19}] // N,
 TableHeadings → {{"Neutral", "DFE1", "DFE2", "DFE3", "DFE4", "DFE5", "DFE6"}, Range[1, 19]}]
```

General: Exp[−852.702] is too small to represent as a normalized machine number; precision may be lost.

General: Exp[−852.702] is too small to represent as a normalized machine number; precision may be lost.

General: Exp[−852.702] is too small to represent as a normalized machine number; precision may be lost.

General: Further output of General::munfl will be suppressed during this calculation.

Out[78]//TableForm=

	1	2	3	4	5	6	7	8	9	1
Neutral	0.161012	0.0964657	0.0746265	0.0634678	0.0565841	0.0518402	0.0483211	0.0455693	0.0433305	0
DFE1	0.184488	0.104069	0.0771662	0.063643	0.0554717	0.0499766	0.0460109	0.0430012	0.0406289	0
DFE2	0.199456	0.109495	0.0794627	0.0644126	0.0553555	0.0492949	0.0449466	0.0416685	0.039104	0
DFE3	0.272676	0.137393	0.0922993	0.0697523	0.0562239	0.0472054	0.0407627	0.035932	0.0321734	0
DFE4	0.210618	0.113657	0.0813119	0.0651207	0.0553911	0.0488923	0.0442396	0.0407409	0.0380113	0
DFE5	0.243946	0.126343	0.0871402	0.0675367	0.0557731	0.0479294	0.0423257	0.038122	0.0348516	0
DFE6	0.205157	0.111612	0.0803968	0.0647641	0.0553643	0.0490811	0.0445788	0.0411896	0.0385424	0

Exponential growth model

In[79]:=

```
TableForm[Table[termsExpPBI[[k]], {k, 1, 7}, {b, 1, 19}] // N,
 TableHeadings → {{"Neutral", "DFE1", "DFE2", "DFE3", "DFE4", "DFE5", "DFE6"}, Range[1, 19]}]
```

Out[79]//TableForm=

	1	2	3	4	5	6	7	8	9	10
Neutral	0.536352	0.120582	0.0621538	0.0427046	0.0330783	0.0272213	0.0232155	0.0202722	0.0180048	0.
DFE1	0.550542	0.122304	0.0611101	0.0411351	0.0315087	0.0257799	0.0219215	0.019114	0.0169638	0.
DFE2	0.553521	0.122749	0.0609239	0.0408132	0.0311752	0.0254682	0.0216388	0.0188594	0.0167342	0.
DFE3	0.557373	0.123376	0.0607031	0.0404019	0.0307418	0.0250599	0.0212667	0.0185234	0.0164305	0.
DFE4	0.551155	0.122393	0.0610707	0.0410685	0.0314402	0.025716	0.0218636	0.0190619	0.0169169	0.
DFE5	0.549016	0.122088	0.0612103	0.0413011	0.031679	0.0259382	0.0220647	0.0192427	0.0170798	0.
DFE6	0.544107	0.121446	0.0615537	0.0418403	0.0322244	0.026442	0.0225184	0.0196496	0.0174459	0.

Nucleotide diversity (expected branch lengths of a sample of 2)

In[80]:=
```
pisPBI = Table[2 × 10^-8 {ejGen10k, ejGenStep, ejGenExp}[[l]][2, bPostBurnIn[[l, k]]], {k, 1, 7}, {l, 1, 3}]; // AbsoluteTiming
```

Out[80]= {104.912, Null}

Nucleotide diversity expected after demographic change using values of B from simulations (obtained after burn-in, but before demographic change)

In[81]:=
```
TableForm[pisPBI // Transpose, TableHeadings → {{"Constant", "Decline", "Expansion"},
    {"Neutral", "DFE1", "DFE2", "DFE3", "DFE4", "DFE5", "DFE6"}}] // N // ScientificForm
```

Out[81]//ScientificForm=

	Neutral	DFE1	DFE2	DFE3	DFE4	DFE5	DFE6
Constant	$4. \times 10^{-4}$	2.864×10^{-4}	2.408×10^{-4}	1.012×10^{-4}	2.16×10^{-4}	1.552×10^{-4}	2.272×10^{-4}
Decline	2.15672×10^{-4}	1.18543×10^{-4}	8.80969×10^{-5}	2.23241×10^{-5}	7.20142×10^{-5}	4.0834×10^{-5}	7.93551×10^{-5}
Expansion	5.19325×10^{-5}	4.61338×10^{-5}	4.49645×10^{-5}	4.34715×10^{-5}	4.5892×10^{-5}	4.67384×10^{-5}	4.87125×10^{-5}

Corresponding final values of B

In[82]:=
```
TableForm[Map[#/#[[1]]&,pisPBI//Transpose],TableHeadings→{{"Constant","Decline","Expansion"},{"Neutral","DFE1","DFE2","DFE3","DFE4","DFE
```

Out[82]//TableForm=

	Neutral	DFE1	DFE2	DFE3	DFE4	DFE5	DFE6
Constant	1.	0.716	0.602	0.253	0.54	0.388	0.568
Decline	1.	0.549643	0.408476	0.103509	0.333906	0.189334	0.367943
Expansion	1.	0.888342	0.865825	0.837078	0.883684	0.899984	0.937996

End

In[83]:=
```
Date[]
```

Out[83]= {2020, 12, 22, 0, 16, 23.453486}