What the heck is an SFS (site frequency spectrum)?

	Site 1
Sample 1	AA
Sample 2	AG
Sample 3	AA
Sample 4	AA
Sample 5	AA
Sample 6	AA
Sample 7	AA
Sample 8	AA

What the heck is an SFS (site frequency spectrum)?

	Site 1
Sample 1	AA
Sample 2	AG
Sample 3	AA
Sample 4	AA
Sample 5	AA
Sample 6	AA
Sample 7	AA
Sample 8	AA
	1

	Site 1	Site 2
Sample 1	AA	CC
Sample 2	AG	CC
Sample 3	AA	CC
Sample 4	AA	CC
Sample 5	AA	CC
Sample 6	AA	CC
Sample 7	AA	CC
Sample 8	AA	CC
	1	0

	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10
Sample 1	AA	CC	TT	AA	AG	TT	TT	CC	TT	AA
Sample 2	AG	CC	CC	AA	AA	TT	TT	CC	TT	AA
Sample 3	AA	CC	TT	AA	AG	TT	TT	CC	TT	AA
Sample 4	AA	CC	TT	AA	AA	TT	TT	CC	TT	GG
Sample 5	AA	CC	CT	AA	AA	TT	TT	CC	TT	AA
Sample 6	AA	CC	TT	AA	AA	TT	TT	CC	TT	AA
Sample 7	AA	CC	TT	AA	AA	TT	CT	CC	TT	AA
Sample 8	AA	CC	TT	AA	AA	TT	TT	CC	TT	AA
	1	0	3	0	2	0	1	0	0	2

A “folded” SFS summarizes the minor allele count (MAC)

Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10
1	0	3	0	2	0	1	0	0	2

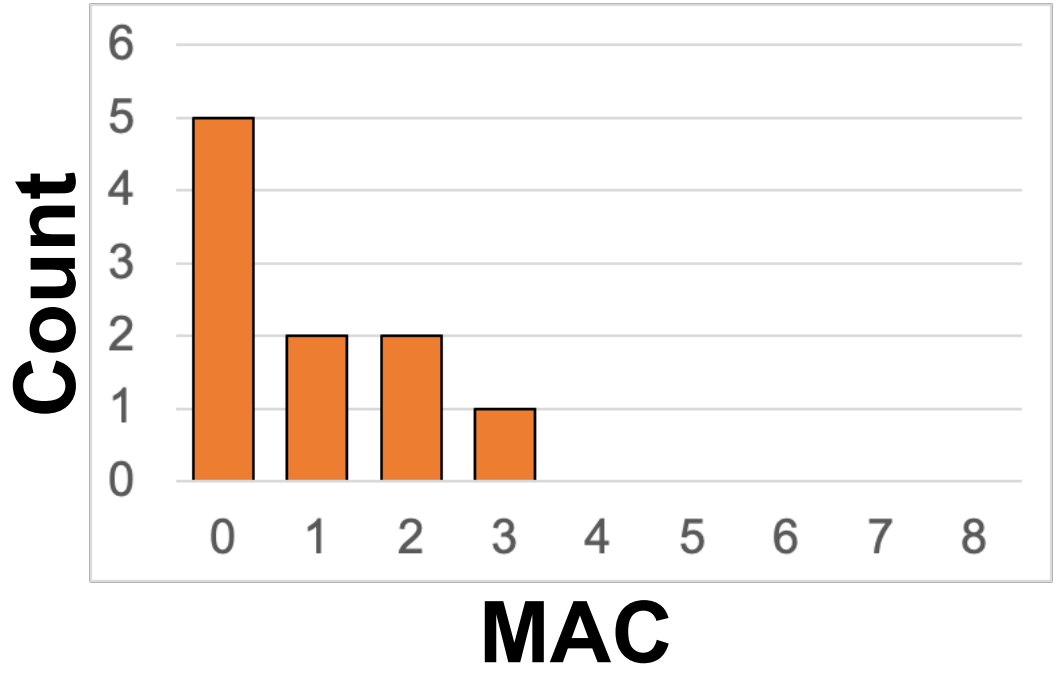
A “folded” SFS summarizes the minor allele count (MAC)

MAC	0	1	2	3	4	5	6	7	8
Count	5	2	2	1	0	0	0	0	0

Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10
1	0	3	0	2	0	1	0	0	2

A “folded” SFS summarizes the minor allele count (MAC)

MAC	0	1	2	3	4	5	6	7	8
Count	5	2	2	1	0	0	0	0	0



An “unfolded” SFS summarizes the derived allele count

Sample 1	AA	CC	TT	AA	AG	TT	TT	CC	TT	AA
Sample 2	AG	CC	CC	AA	AA	TT	TT	CC	TT	AA
Sample 3	AA	CC	TT	AA	AG	TT	TT	CC	TT	AA
Sample 4	AA	CC	TT	AA	AA	TT	TT	CC	TT	GG
Sample 5	AA	CC	CT	AA	AA	TT	TT	CC	TT	AA
Sample 6	AA	CC	TT	AA	AA	TT	TT	CC	TT	AA
Sample 7	AA	CC	TT	AA	AA	TT	CT	CC	TT	AA
Sample 8	AA	CC	TT	AA	AA	TT	TT	CC	TT	AA
	1	0	3	0	2	0	1	0	0	2

	A	C	C	A	A	C	T	C	T	A/G
Sample 1	AA	CC	TT	AA	AG	TT	TT	CC	TT	AA
Sample 2	AG	CC	CC	AA	AA	TT	TT	CC	TT	AA
Sample 3	AA	CC	TT	AA	AG	TT	TT	CC	TT	AA
Sample 4	AA	CC	TT	AA	AA	TT	TT	CC	TT	GG
Sample 5	AA	CC	CT	AA	AA	TT	TT	CC	TT	AA
Sample 6	AA	CC	TT	AA	AA	TT	TT	CC	TT	AA
Sample 7	AA	CC	TT	AA	AA	TT	CT	CC	TT	AA
Sample 8	AA	CC	TT	AA	AA	TT	TT	CC	TT	AA
	1	0	3	0	2	0	1	0	0	2

	A	C	C	A	A	C	T	C	T	A/G
Sample 1	AA	CC	TT	AA	AG	TT	TT	CC	TT	AA
Sample 2	AG	CC	CC	AA	AA	TT	TT	CC	TT	AA
Sample 3	AA	CC	TT	AA	AG	TT	TT	CC	TT	AA
Sample 4	AA	CC	TT	AA	AA	TT	TT	CC	TT	GG
Sample 5	AA	CC	CT	AA	AA	TT	TT	CC	TT	AA
Sample 6	AA	CC	TT	AA	AA	TT	TT	CC	TT	AA
Sample 7	AA	CC	TT	AA	AA	TT	CT	CC	TT	AA
Sample 8	AA	CC	TT	AA	AA	TT	TT	CC	TT	AA
	1	0	13	0	2	16	1	0	0	–

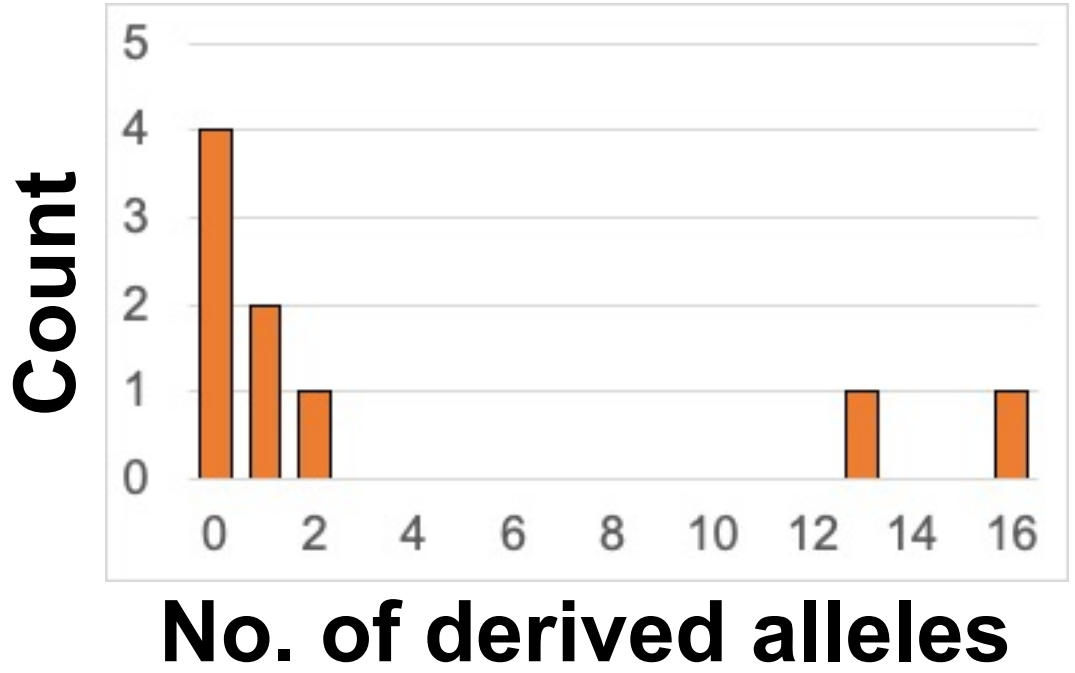
An “unfolded” SFS summarizes the derived allele count

MAC	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Count	4	2	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1

Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10
1	0	13	0	2	16	1	0	0	–

An “unfolded” SFS summarizes the derived allele count

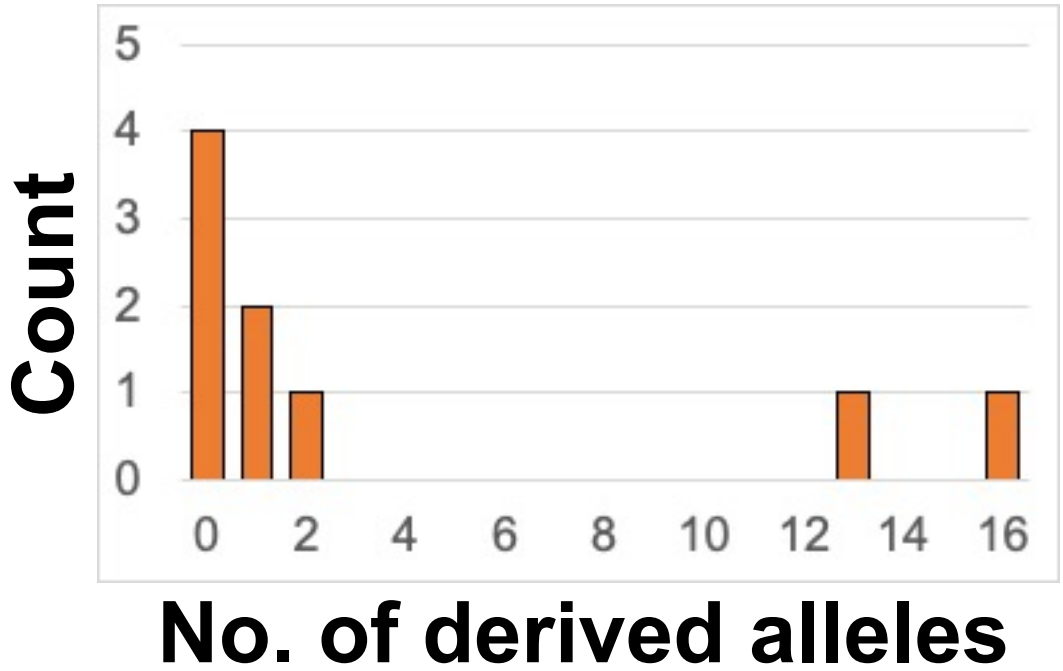
MAC	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Count	4	2	1	0	0	0	0	0	0	0	0	0	0	1	0	0	1

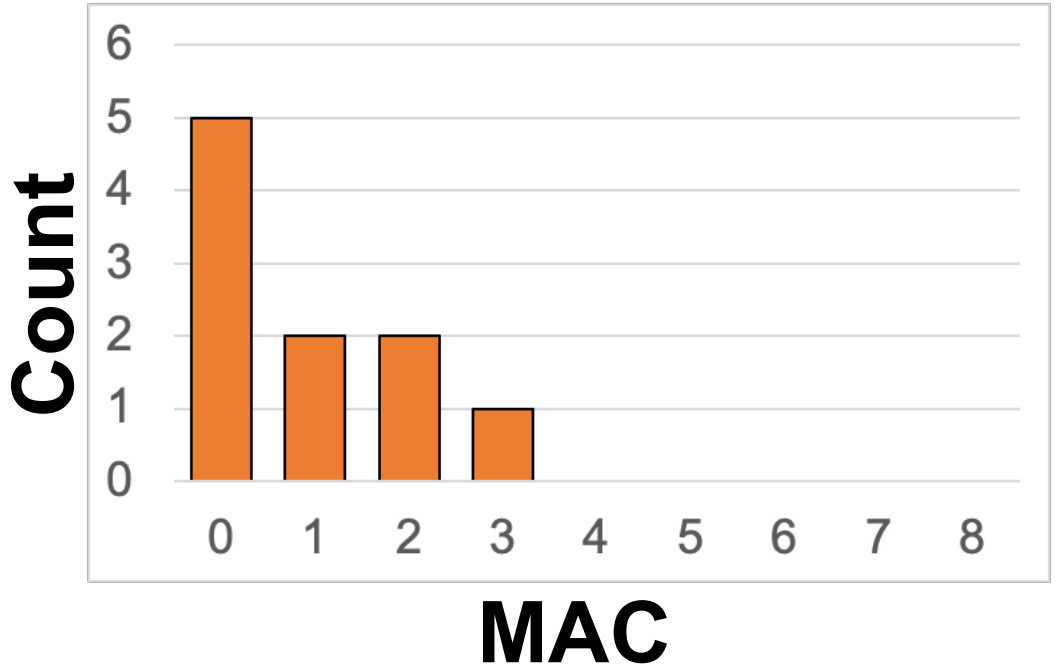


Unfolded vs folded SFS

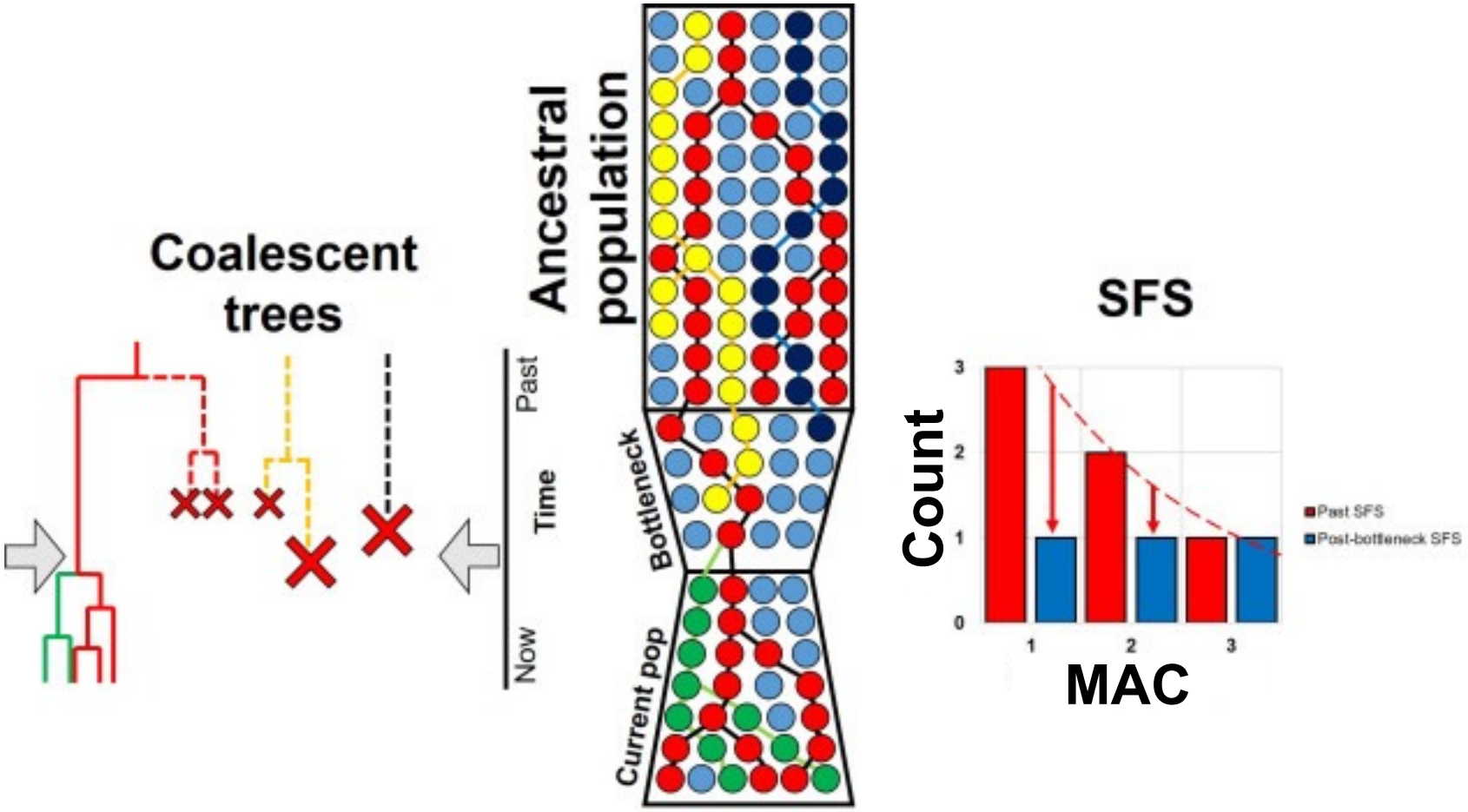
- *2N* entries
- More information (more accurate inferences)
- Requires data from outgroup to estimate ancestral state

- *N* entries
- Less information (less accurate inferences)
- Only requires data from the species of interest



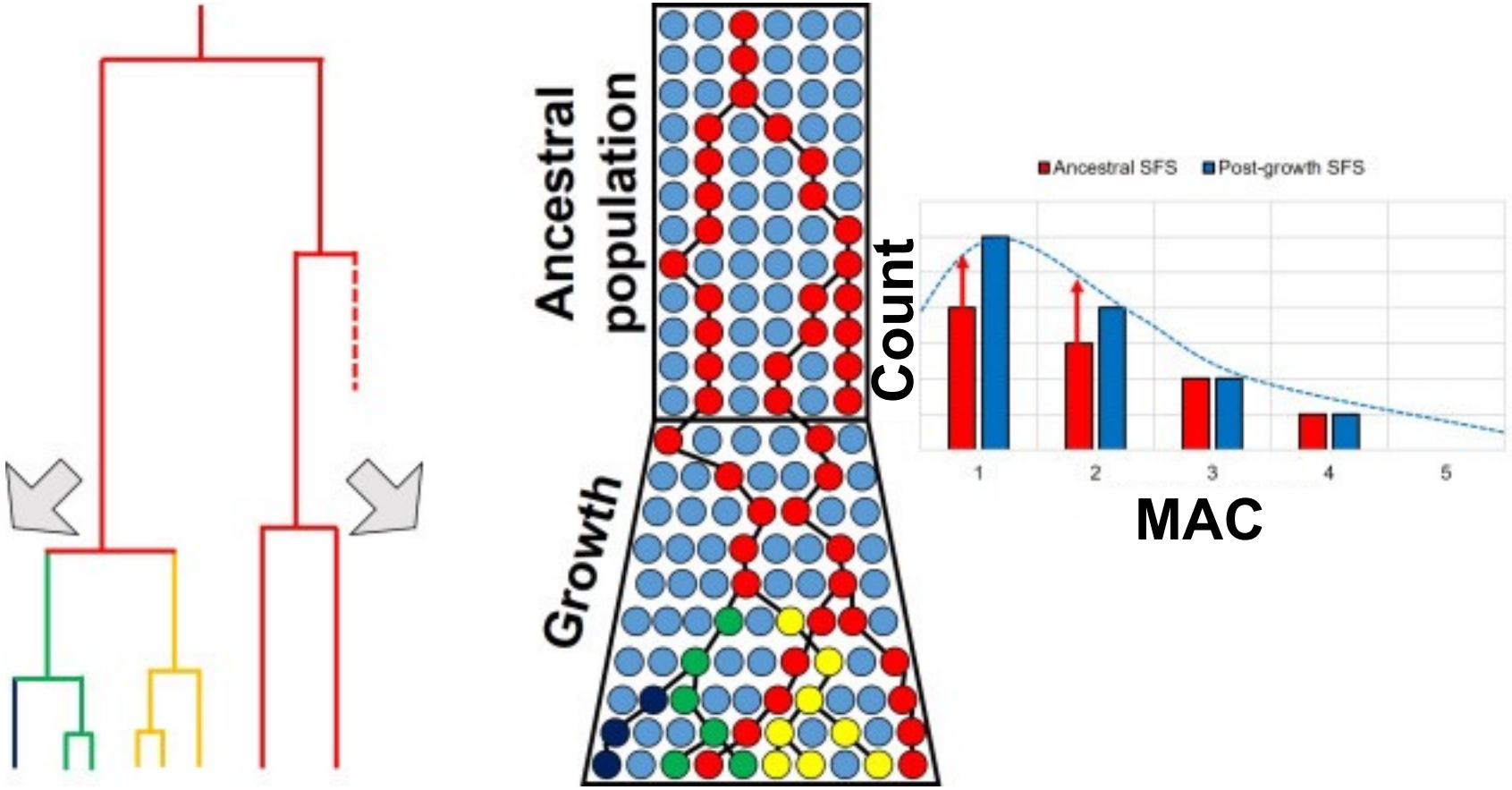


Demography impacts the SFS



A bottleneck “evens” out the counts of different minor allele counts (e.g. reduces the number of singletons)

Demography impacts the SFS



Growth/an expansion leads to an excess of singleton sites (a more uneven site frequency spectrum)

Can use the SFS to infer demography

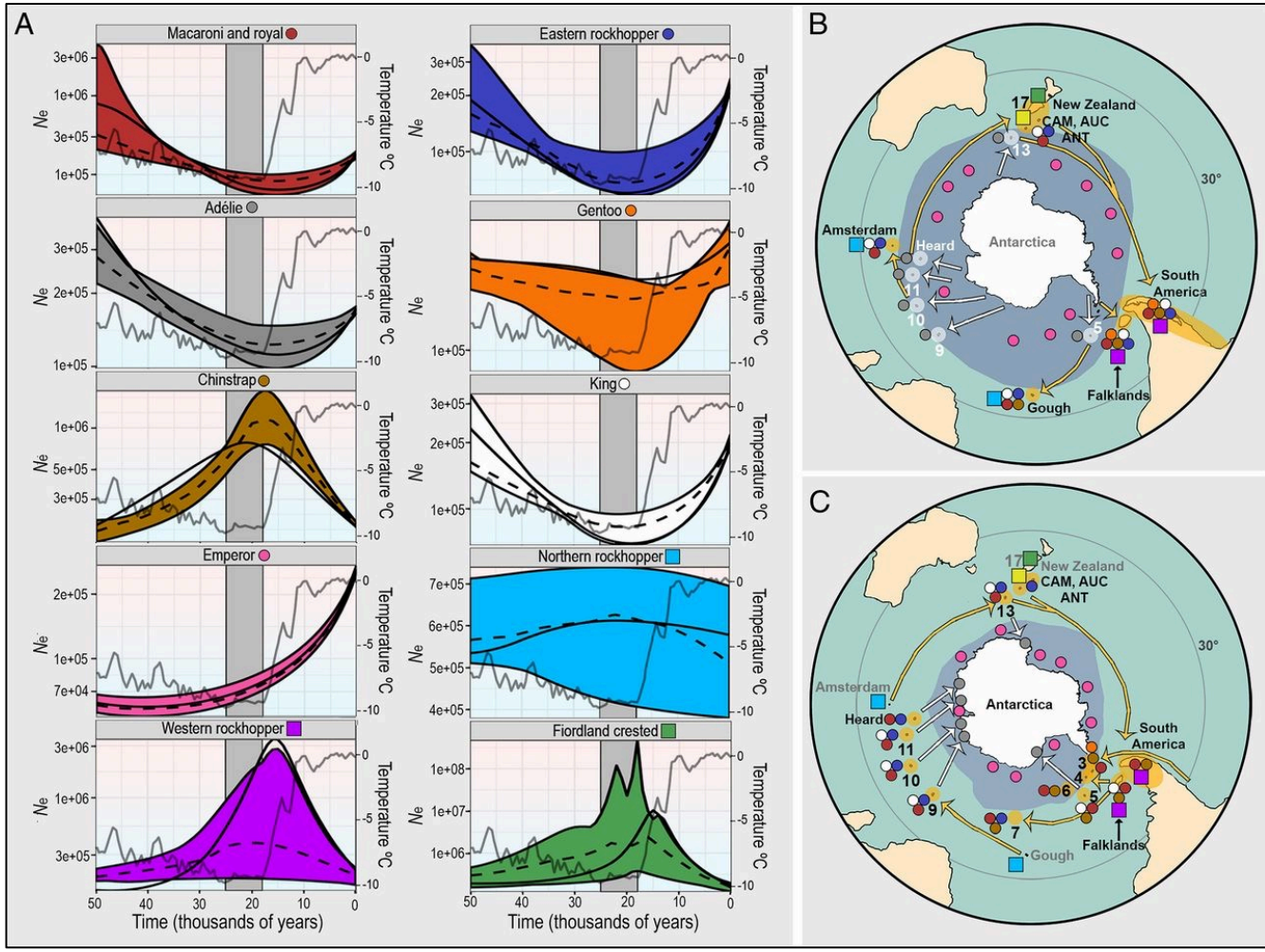
Bunch of approaches for doing this, including:

- Fastsimcoal2
- Dadi
- CubSFS

Can use the SFS to infer demography

Bunch of approaches for doing this, including:

- Fastsimcoal2
- Dadi
- CubSFS



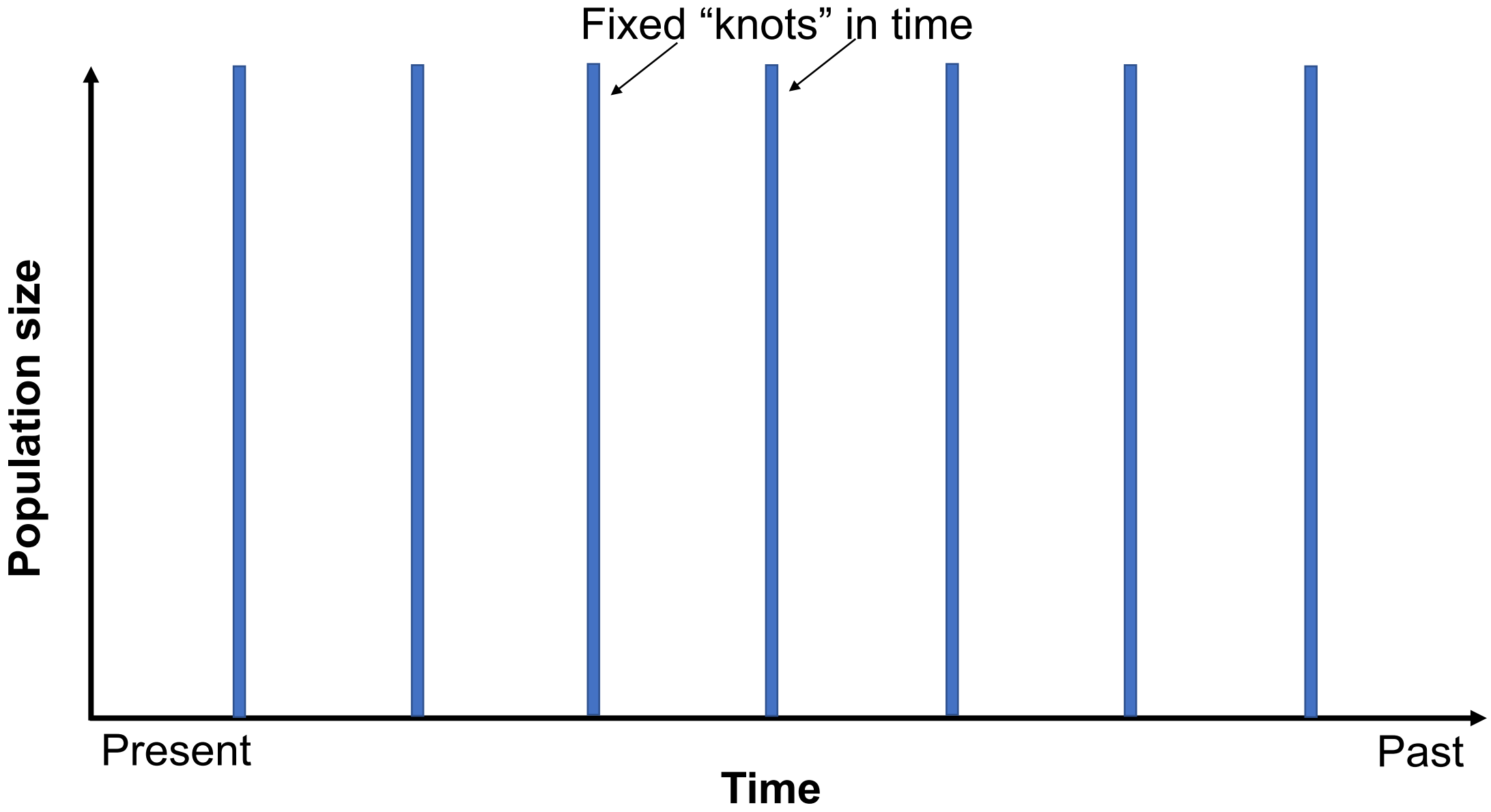
Receding ice drove parallel expansions in Southern Ocean penguins

Theresa L. Cole[a,b,1], Ludovic Dutoit[a,2], Nicolas Dussex[c,d,2], Tom Hart[e,2], Alana Alexander[d,2], Jane L. Younger[f], Gemma V. Clucas[g,h], María José Frugone[i,j], Yves Cherel[k], Richard Cuthbert[l,m], Ursula Ellenberg[n,o], Steven R. Fiddaman[p], Johanna Hiscock[q], David Houston[r], Pierre Jouventin[s], Thomas Mattern[a], Gary Miller[t,u], Colin Miskelly[v], Paul Nolan[w], Michael J. Polito[x], Petra Quillfeldt[y], Peter G. Ryan[z], Adrian Smith[p], Alan J. D. Tennyson[v], David Thompson[aa], Barbara Wienecke[bb], Juliana A. Vianna[cc], and Jonathan M. Waters[a]

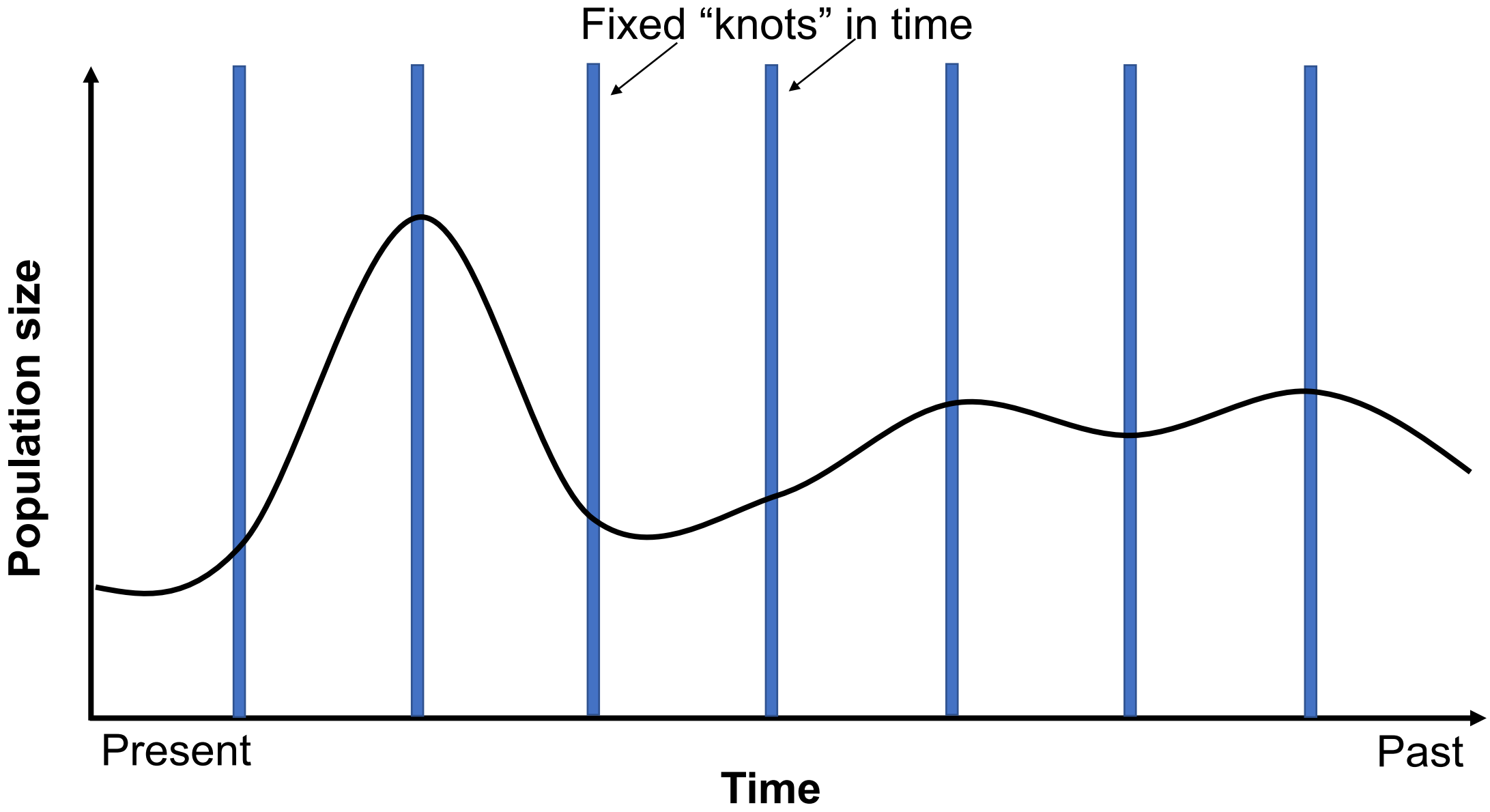
26690–26696 | PNAS | December 26, 2019 | vol. 116 | no. 52

www.pnas.org/cgi/doi/10.1073/pnas.1904048116

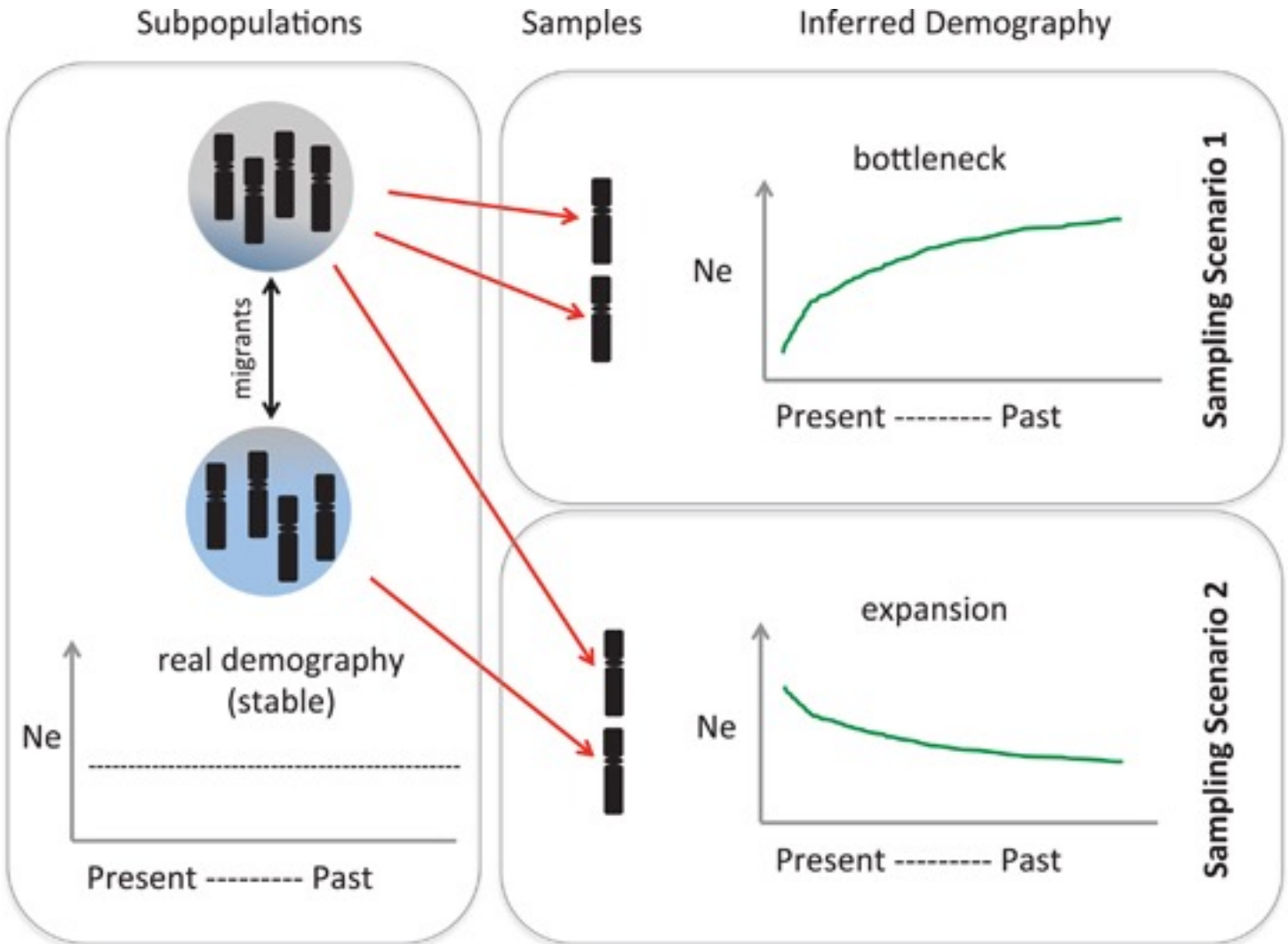
CubSFS is quick...



CubSFS is quick...



Like many models, only valid if no population structure



Using CubSFS to investigate historical demography of kanakana (lamprey)

Using CubSFS to investigate historical demography of kanakana (lamprey)

Access to folded SFS data thanks to Allison Miller:
https://gemmell-lab.otago.ac.nz/our-team/21-team/phd-students/153-allison-miller

Lamprey community science:
https://www.inaturalist.org/projects/lamps_for_champs_obs
https://a3miller.wixsite.com/fishybites