GENETICS

Supporting Information
http://www.genetics.org/cgi/content/full/genetics.110.123059/DC1

Genomic Differentiation Between Temperate and Tropical Australian Populations of *Drosophila melanogaster*

Bryan Kolaczkowski, Andrew D. Kern, Alisha K. Holloway and David J. Begun

Copyright © 2011 by the Genetics Society of America
DOI: 10.1534/genetics.110.123059

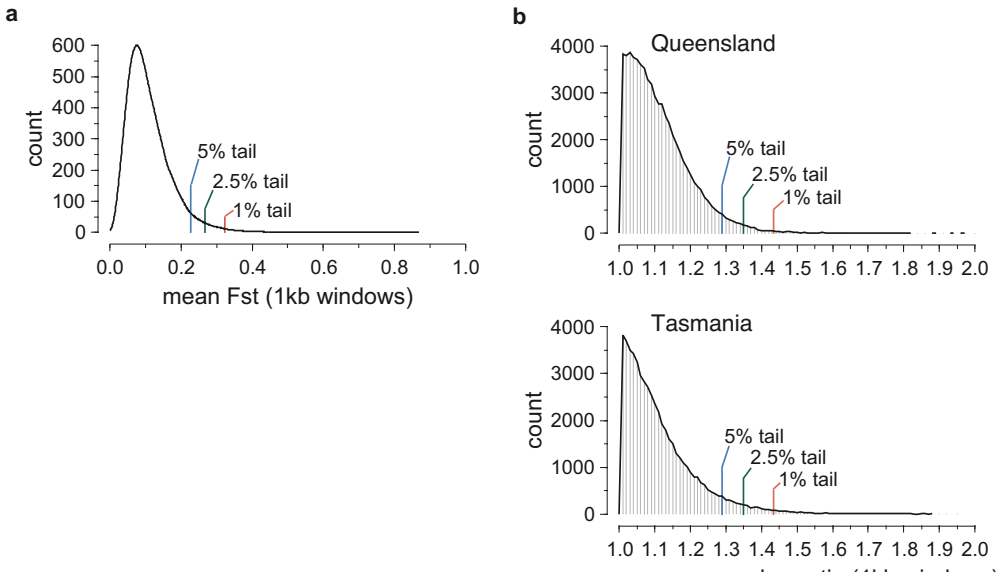

FIGURE S1.—Upper-tail F_{ST} and copy-number ratio cutoffs used in this study. We bin nonoverlapping 1kb genomic windows of F_{ST} (a) and copy-number ratio (b) and plot the number of windows in each bin. Tail cutoffs of 1, 2.5 and 5% are indicated in each panel.

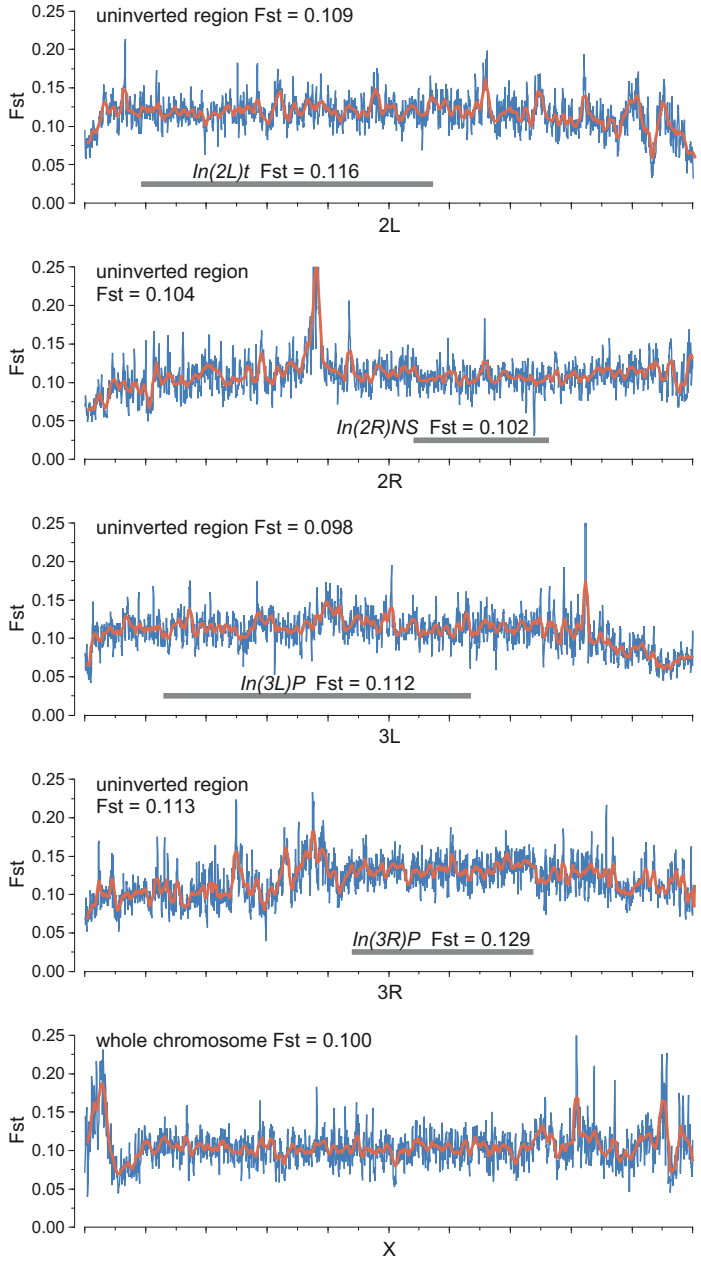

FIGURE S2.—We plot F_{ST} for standard and inverted regions of each chromosome arm. Inverted regions are indicated by gray horizontal lines. Blue series indicate average F_{ST} values over 25kb windows slid every 10kb; red lines show 200kb windows slid 50kb at a time. Overall F_{ST} across each region (standardvs. inverted) is also indicated in each panel. Note that there is no inversion on the X chromosome.

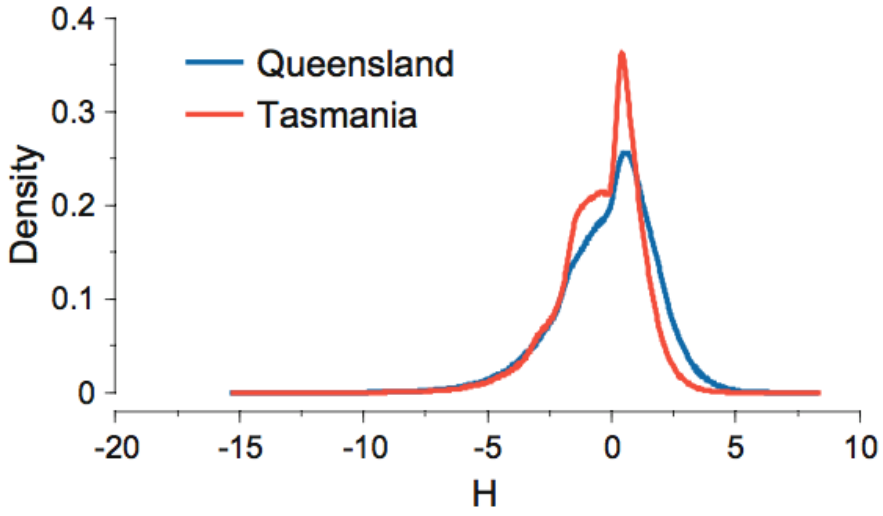

FIGURE S3.—A comparison of the distributions of H statistics computed in 1kb windows in both the Queensland and the Tasmanian samples. This is a modified version of Fay and Wu's H which exludes singleton variants. See main text for details. The medians of these distributions are significantly different from one another in a Wilcoxon rank sum test ($p < 2.2 \times 10^{-16}$).

FIGURE S4.—Fixed differences between Queensland and Tasmania at the *Irc* gene radically alter pre-mRNA structure. Computationally-inferred secondary structure of *Irc* pre-mRNA is shown for Queensland and Tasmania alleles.

TABLE S1

We report average polymorphism per 1kb (π) and Fay and Wu's *H* for each population, as well as mean F_{ST} and size of highly-differentiated genomic regions for the normally-recombining portion of each chromosome arm.

	π		Fay and Wu's H			differentiated
chromosome	Queensland	Tasmania	Queensland	Tasmania	F_{ST}	region size
2L	5.060	3.146	-0.108	-0.500	0.116	2206
2R	4.905	3.284	-0.272	-0.428	0.107	2507
3L	5.051	3.130	-0.125	-0.426	0.111	2013
3R	5.010	3.117	0.043	-0.442	0.124	3295
X	2.926	2.121	-0.500	-0.456	0.097	1899

TABLE S2

Genes in the top 2.5% of 3'UTR differentiation.

Gene ID	3'UTR F_{ST}	Gene Name	Gene ID	3'UTR F_{ST}	Gene Name
FBgn0038783	0.692282	CG4367	FBgn0035028	0.341374	Start1
FBgn0038827	0.66719	Fancd2	FBgn0259680	0.337828	Pkcdelta
FBgn0033809	0.620316	CG4630	FBgn0031681	0.337809	pgant5
FBgn0033808	0.620316	CG4627	FBgn0031861	0.335605	CG17375
FBgn0038225	0.596327	soti	FBgn0051320	0.334101	CG31320
FBgn0031491	0.561771	alpha4GT1	FBgn0038223	0.334016	CG8538
FBgn0038652	0.544105	CG7720	FBgn0259984	0.333621	kuz
FBgn0032974	0.532426	CG3651	FBgn0034820	0.331211	CG13538
FBgn0039396	0.523533	CcapR	FBgn0011289	0.327801	TfIIA-L
FBgn0038478	0.521974	cal1	FBgn0016054	0.32776	phr6-4
FBgn0038292	0.515821	CG3987	FBgn0024832	0.326232	AP-50
FBgn0032409	0.469866	Ced-12	FBgn0019968	0.32613	Khc-73
FBgn0032520	0.468198	CG10859	FBgn0031619	0.321426	CG3355
FBgn0032318	0.458057	CG14072	FBgn0085374	0.320338	CG34345
FBgn0010441	0.446403	pll	FBgn0032515	0.319843	loqs
FBgn0038124	0.436888	CG14380	FBgn0031752	0.317964	CG9044
FBgn0032465	0.430473	CG12404	FBgn0024432	0.317639	Dlc90F
FBgn0023415	0.426255	Acp32CD	FBgn0053092	0.317186	CG33092
FBgn0031284	0.424294	CG3876	FBgn0029161	0.316588	slmo
FBgn0039049	0.421011	CG6726	FBgn0085208	0.316588	CG34179
FBgn0020368	0.419943	Vha68-1	FBgn0035370	0.315363	CG1240
FBgn0043471	0.418173	kappaTry	FBgn0033128	0.315135	Tsp42Eg
FBgn0004888	0.41782	Scsalpha	FBgn0046776	0.31513	CG14033
FBgn0042710	0.414471	Hex-t2	FBgn0050050	0.314525	CG30050
FBgn0010222	0.399271	Nmdmc	FBgn0011244	0.313799	Hsp60B
FBgn0024947	0.395613	NTPase	FBgn0259966	0.311797	Sfp51E
FBgn0038819	0.393874	Cpr92F	FBgn0028482	0.3112	CG16857
FBgn0039061	0.384126	Ir	FBgn0050043	0.307662	CG30043
FBgn0050021	0.374556	skf	FBgn0033304	0.307415	Cyp6a13
FBgn0039293	0.369886	CG11851	FBgn0039969	0.306813	Fis1
FBgn0003248	0.36957	Rh2	FBgn0032474	0.302561	DnaJ-H
FBgn0026576	0.362997	CG5991	FBgn0039107	0.301822	CG10300
FBgn0014141	0.362454	cher	FBgn0040397	0.30087	CG3655
FBgn0031529	0.36213	CG9662	FBgn0020767	0.300498	Spred
FBgn0026758	0.356538	Trf2	FBgn0033806	0.300339	CG4616
FBgn0038815	0.354985	CG5466	FBgn0022073	0.299715	Thor
FBgn0086677	0.354284	jeb	FBgn0032763	0.297223	CG17568
FBgn0034263	0.353819	CG10934	_`Bgn0013433	0.296955	beat-Ia
FBgn0039238	0.352161	CG7016	FBgn0033252	0.296132	CG12769

FBgn0039239	0.352161	CG13641	FBgn0038407	0.296092	CG6126
FBgn0031522	0.344939	CG3285	FBgn0015844	0.295943	Xpd

We calculated the mean F_{ST} of each gene's 3'UTR region and list the top 2.5%.

TABLE S3

Genes in the top 2.5% of whole-gene nonsynonymous F_{ST} differentiation.

Gene ID	Nonsyn F_{ST}	Gene Name	Gene ID	Nonsyn F_{ST}	Gene Name
FBgn0031345	0.440546	CG18132	FBgn0035687	0.261204	CG13296
FBgn0086677	0.432406	jeb	FBgn0051380	0.260273	CG31380
FBgn0034288	0.41522	CG5084	FBgn0035216	0.257985	CG9168
FBgn0032331	0.401399	CG14913	FBgn0031321	0.255153	Tgt
FBgn0033093	0.38458	CG3270	FBgn0036951	0.255088	CG7017
FBgn0033697	0.384211	Cyp6t3	FBgn0039413	0.254965	CG14556
FBgn0053339	0.381734	CG33339	FBgn0025676	0.254259	CkIIalpha-i3
FBgn0053205	0.376138	CG33205	FBgn0260011	0.254035	nimC4
FBgn0035276	0.37592	CG12022	FBgn0021764	0.253939	sdk
FBgn0029173	0.373758	fu2	FBgn0038195	0.253804	CG3061
FBgn0051082	0.372544	CG31082	FBgn0003486	0.2534	spo
FBgn0025621	0.372522	CG16989	FBgn0036593	0.253169	CG13048
FBgn0085401	0.367414	CG34372	FBgn0051265	0.252429	CG31265
FBgn0011285	0.352576	S6kII	FBgn0035268	0.251505	CG8001
FBgn0051851	0.338734	CG31851	FBgn0031975	0.248371	Tg
FBgn0066101	0.331073	LpR1	FBgn0052104	0.248324	CG32104
FBgn0045476	0.329101	Gr64e	FBgn0039167	0.248287	CG17786
FBgn0015905	0.328591	ast	FBgn0031739	0.247517	CG14005
FBgn0035313	0.328007	CG13810	FBgn0032843	0.247437	CG10730
FBgn0085449	0.325412	CG34420	FBgn0015230	0.246036	Glut3
FBgn0003861	0.324953	trp	FBgn0034295	0.245982	CG10911
FBgn0051205	0.319649	CG31205	FBgn0052154	0.245111	CG32154
FBgn0039183	0.316728	Dis3	FBgn0033473	0.244758	CG12128
FBgn0038144	0.316311	CG8870	FBgn0028482	0.244381	CG16857
FBgn0052549	0.31354	CG32549	FBgn0033742	0.244029	CG8550
FBgn0038153	0.311496	Ir87a	FBgn0031782	0.243645	WDR79
FBgn0039528	0.310709	dsd	FBgn0039079	0.243057	Ir94g
FBgn0036541	0.306629	CG12486	FBgn0036320	0.242916	CG10943
FBgn0040045	0.305202	CG12460	FBgn0028852	0.242757	CG15262
FBgn0038139	0.304808	CG8795	FBgn0038125	0.242423	CG8141
FBgn0031195	0.304405	CG17600	FBgn0039201	0.242182	CG13617
FBgn0065035	0.30298	AlkB	FBgn0037989	0.24141	CG14741
FBgn0034513	0.296861	CG13423	FBgn0043796	0.241249	CG12219
FBgn0038550	0.296205	CG17801	FBgn0035918	0.241212	Cdc6
FBgn0036249	0.294651	CG11560	FBgn0037817	0.240972	Cyp12e1
FBgn0020272	0.29456	mst	FBgn0034422	0.240814	CG7137
FBgn0032045	0.291513	CG13087	FBgn0036503	0.240784	CG13454
FBgn0085213	0.290632	CG34184	FBgn0036952	0.239612	CG6933
FBgn0052751	0.290399	CG3275[illegible]	FBgn0039067	0.23927	wda

FBgn0038238	0.290022	CG14854	FBgn0085481	0.238273	CG34452
FBgn0032066	0.289623	CG9463	FBgn0043005	0.23733	CG10251
FBgn0053503	0.288901	Cyp12d1-d	FBgn0039227	0.237257	polybromo
FBgn0041607	0.288162	asparagine-synthetase	FBgn0045487	0.236246	Gr36a
FBgn0260466	0.287405	Indy-2	FBgn0260873	0.236107	CG42583
FBgn0026737	0.286948	CG6171	FBgn0032370	0.236036	CG12307
FBgn0051005	0.286607	CG31005	FBgn0020640	0.235496	Lcp65Ae
FBgn0053012	0.285977	CG33012	FBgn0010328	0.234114	woc
FBgn0030946	0.284422	CG6659	FBgn0085253	0.233799	CG34224
FBgn0085305	0.282863	CG34276	FBgn0030054	0.233653	Caf1-180
FBgn0030752	0.282683	CG9947	FBgn0038133	0.232298	Osi22
FBgn0038850	0.280486	CG17279	FBgn0039684	0.231514	Obp99d
FBgn0039246	0.279792	CG10845	FBgn0030033	0.230891	CG1387
FBgn0033599	0.279459	CG13223	FBgn0033443	0.230715	CG1698
FBgn0036948	0.277711	CG7298	FBgn0039343	0.230581	CG5111
FBgn0259199	0.277513	CG42303	FBgn0035970	0.230459	CG4483
FBgn0025866	0.277104	CalpB	FBgn0039467	0.230335	CG14253
FBgn0039083	0.276633	CG10177	FBgn0035771	0.229884	sec63
FBgn0032803	0.275137	CG13082	FBgn0038912	0.229665	CG6656
FBgn0033186	0.275072	CG1602	FBgn0033698	0.229018	CG8858
FBgn0040391	0.273132	CG2854	FBgn0014469	0.228657	Cyp4e2
FBgn0052107	0.271473	CG32107	FBgn0031961	0.227479	CG7102
FBgn0033187	0.27141	CG2144	FBgn0046689	0.227113	Takl1
FBgn0051251	0.268618	CG31251	FBgn0054041	0.226485	CG34041
FBgn0051461	0.267248	CG31461	FBgn0034141	0.22639	CG8311
FBgn0085341	0.265634	CG34312	FBgn0033850	0.226218	CG13331
FBgn0036193	0.265447	CG14135	FBgn0032142	0.225804	CG13120
FBgn0051099	0.265062	CG31099	FBgn0037533	0.225779	CD98hc
FBgn0036062	0.264058	CG6685	FBgn0039080	0.225681	Ir94h
FBgn0032484	0.263439	kek4	FBgn0031817	0.225339	CG9531
FBgn0013949	0.262919	Ela	FBgn0037248	0.225186	Spargel
FBgn0033289	0.262138	CG2121	FBgn0031429	0.224589	CG15393
FBgn0032602	0.26173	CG13278			

We calculated the mean nonsynonymous F_{ST} of each gene and list the top 2.5%

TABLE S4

Genes in the top 2.5% of individual-domain nonsynonymous F_{ST} differentiation.

Gene ID	Gene Name	Domain	Nonsyn F_{ST}
FBgn0033093	CG3270	DAO	0.38458
FBgn0038260	CG14855	MFS 1	0.384254
FBgn0033697	Cyp6t3	p450	0.384211
FBgn0034295	CG10911	DUF725	0.360936
FBgn0051380	CG31380	APH	0.342056
FBgn0003510	Sry-alpha	Serendipity A	0.33336
FBgn0045476	Gr64e	7tm 7	0.329101
FBgn0051205	CG31205	Trypsin	0.319649
FBgn0032602	CG13278	ASC	0.298351
FBgn0034513	CG13423	Peptidase C1 2	0.296861
FBgn0085213	CG34184	DM4 12	0.290632
FBgn0260466	Indy-2	Na sulph symp	0.287405
FBgn0030946	CG6659	Dpy19	0.284422
FBgn0030752	CG9947	CDC50	0.282683
FBgn0038850	CG17279	JHBP	0.280486
FBgn0028852	CG15262	NOT2 3 5	0.278706
FBgn0020377	Sr-CII	MAM	0.27817
FBgn0031305	Iris	DUF3610	0.27248
FBgn0033443	CG1698	SNF	0.268159
FBgn0051099	CG31099	DUF227	0.265062
FBgn0021764	sdk	fn3	0.263057
FBgn0032066	CG9463	Glyco hydro 38C	0.258052
FBgn0038005	Cyp313a5	p450	0.257784
FBgn0003486	spo	p450	0.2534
FBgn0038465	Irc	An peroxidase	0.247982
FBgn0054005	CG34005	DUF725	0.242942
FBgn0038541	TyrRII	7tm 1	0.239625
FBgn0054049	CG34049	CAP	0.237954
FBgn0045487	Gr36a	7tm 7	0.236246

We calculated the mean nonsynonymous F_{ST} of each PFam domain in each gene and took the most-differentiated domain as that gene's representative domain. We list list the top 2.5% of individual-domain differentiated genes.

TABLE S5

Significantly-enriched Gene Ontology categories in top 2.5% 1kb F_{ST} regions.

Biological Process		
GO accession	P-value	Description
GO:0007424	2.49E-09	open tracheal system development
GO:0007428	1.56E-06	primary branching, open tracheal system
GO:0007165	1.56E-06	signal transduction
GO:0007427	1.56E-06	epithelial cell migration, open tracheal system
GO:0002121	2.21E-06	inter-male aggressive behavior
GO:0007509	2.62E-06	mesoderm migration
GO:0042051	2.62E-06	compound eye photoreceptor development
GO:0006355	3.91E-06	regulation of transcription, DNA-dependent
GO:0007156	1.37E-05	homophilic cell adhesion
GO:0048477	1.78E-05	oogenesis
GO:0007435	1.78E-05	salivary gland morphogenesis
GO:0007507	1.82E-05	heart development
GO:0008543	2.13E-05	fibroblast growth factor receptor signaling pathway
GO:0035152	4.50E-05	regulation of tube architecture, open tracheal system
GO:0007614	8.48E-05	short-term memory
GO:0006816	8.48E-05	calcium ion transport
GO:0007431	1.54E-04	salivary gland development
GO:0007155	1.70E-04	cell adhesion
GO:0016044	2.50E-04	membrane organization
GO:0007293	4.51E-04	germarium-derived egg chamber formation
GO:0007411	4.51E-04	axon guidance
GO:0042048	4.51E-04	olfactory behavior
GO:0030707	4.51E-04	ovarian follicle cell development
GO:0008101	4.87E-04	decapentaplegic receptor signaling pathway
GO:0045570	4.87E-04	regulation of imaginal disc growth
GO:0035172	4.87E-04	hemocyte proliferation
GO:0007443	4.88E-04	Malpighian tubule morphogenesis
GO:0048813	4.88E-04	dendrite morphogenesis
GO:0048190	5.67E-04	wing disc dorsal/ventral pattern formation
GO:0007422	7.20E-04	peripheral nervous system development
GO:0016055	7.20E-04	Wnt receptor signaling pathway
GO:0007611	7.20E-04	learning or memory
GO:0007426	1.12E-03	tracheal outgrowth, open tracheal system
GO:0006813	1.21E-03	potassium ion transport
GO:0007517	1.32E-03	muscle organ development
GO:0048066	1.52E-03	pigmentation during development
GO:0007297	1.52E-03	ovarian follicle cell migration
GO:0048675	1.52E-03	axon extension

GO:0006811	1.67E-03	ion transport
GO:0007476	1.68E-03	imaginal disc-derived wing morphogenesis
GO:0030718	1.72E-03	germ-line stem cell maintenance
GO:0007619	2.08E-03	courtship behavior
GO:0045449	2.19E-03	regulation of transcription
GO:0007379	2.38E-03	segment specification
GO:0007417	3.00E-03	central nervous system development
GO:0007399	3.00E-03	nervous system development
GO:0030097	3.15E-03	hemopoiesis
GO:0007274	3.15E-03	neuromuscular synaptic transmission
GO:0007265	3.15E-03	Ras protein signal transduction
GO:0042078	3.15E-03	germ-line stem cell division
GO:0016199	3.46E-03	axon midline choice point recognition
GO:0007602	3.46E-03	phototransduction
GO:0048666	3.54E-03	neuron development
GO:0008355	3.91E-03	olfactory learning
GO:0008407	4.18E-03	bristle morphogenesis
GO:0016477	4.18E-03	cell migration
GO:0016339	4.96E-03	calcium-dependent cell-cell adhesion
GO:0055085	5.47E-03	transmembrane transport
GO:0008063	5.77E-03	Toll signaling pathway
GO:0008354	5.77E-03	germ cell migration
GO:0006357	6.28E-03	regulation of transcription from RNA polymerase II promoter
GO:0035147	6.60E-03	branch fusion, open tracheal system
GO:0008344	6.60E-03	adult locomotory behavior
GO:0009953	6.60E-03	dorsal/ventral pattern formation
GO:0035277	6.60E-03	spiracle morphogenesis, open tracheal system
GO:0006325	6.93E-03	chromatin organization
GO:0007494	6.93E-03	midgut development
GO:0002009	6.93E-03	morphogenesis of an epithelium
GO:0006468	8.66E-03	protein amino acid phosphorylation
GO:0019991	8.81E-03	septate junction assembly
GO:0007442	8.81E-03	hindgut morphogenesis
GO:0007291	8.81E-03	sperm individualization
GO:0007294	8.81E-03	germarium-derived oocyte fate determination
GO:0035071	9.76E-03	salivary gland cell autophagic cell death
GO:0007298	9.76E-03	border follicle cell migration
GO:0008104	1.11E-02	protein localization
GO:0000381	1.11E-02	regulation of alternative nuclear mRNA splicing, via spliceosome
GO:0017148	1.11E-02	negative regulation of translation
GO:0051225	1.11E-02	spindle assembly

GO:0001745	1.11E-02	compound eye morphogenesis
GO:0008360	1.12E-02	regulation of cell shape
GO:0007391	1.18E-02	dorsal closure
GO:0007498	1.35E-02	mesoderm development
GO:0007179	1.42E-02	transforming growth factor beta receptor signaling pathway
GO:0007350	1.42E-02	blastoderm segmentation
GO:0016481	1.45E-02	negative regulation of transcription
GO:0001700	1.49E-02	embryonic development via the syncytial blastoderm
GO:0007409	1.79E-02	axonogenesis
GO:0030162	1.80E-02	regulation of proteolysis
GO:0048749	1.95E-02	compound eye development
GO:0007018	1.95E-02	microtubule-based movement
GO:0007268	1.98E-02	synaptic transmission
GO:0007275	2.04E-02	multicellular organismal development
GO:0035023	2.20E-02	regulation of Rho protein signal transduction
GO:0007367	2.20E-02	segment polarity determination
GO:0006911	2.50E-02	phagocytosis, engulfment
GO:0048102	2.57E-02	autophagic cell death
GO:0009987	2.72E-02	cellular process
GO:0006096	2.82E-02	glycolysis
GO:0016318	3.31E-02	ommatidial rotation
GO:0008045	3.31E-02	motor axon guidance
GO:0030036	3.80E-02	actin cytoskeleton organization
GO:0046843	3.82E-02	dorsal appendage formation
GO:0045475	3.82E-02	locomotor rhythm
GO:0006342	3.82E-02	chromatin silencing
GO:0051726	3.82E-02	regulation of cell cycle
GO:0007474	3.82E-02	imaginal disc-derived wing vein specification
GO:0006350	3.82E-02	transcription
GO:0007186	4.03E-02	G-protein coupled receptor protein signaling pathway
GO:0000122	4.78E-02	negative regulation of transcription from RNA polymerase II promoter

Molecular Function

GO accession	P-value	Description
GO:0005515	1.57E-08	protein binding
GO:0003700	8.12E-07	transcription factor activity
GO:0005509	1.93E-04	calcium ion binding
GO:0004889	3.64E-04	nicotinic acetylcholine-activated cation-selective channel activity
GO:0003729	3.64E-04	mRNA binding
GO:0003702	3.64E-04	RNA polymerase II transcription factor activity
GO:0004871	3.64E-04	signal transducer activity

GO:0043565	8.02E-04	sequence-specific DNA binding
GO:0008188	1.87E-03	neuropeptide receptor activity
GO:0003704	1.89E-03	specific RNA polymerase II transcription factor activity
GO:0003777	3.08E-03	microtubule motor activity
GO:0005096	5.38E-03	GTPase activator activity
GO:0016566	5.92E-03	specific transcriptional repressor activity
GO:0016563	6.18E-03	transcription activator activity
GO:0005249	7.73E-03	voltage-gated potassium channel activity
GO:0003723	8.11E-03	RNA binding
GO:0003779	9.12E-03	actin binding
GO:0004888	1.02E-02	transmembrane receptor activity
GO:0005085	1.02E-02	guanyl-nucleotide exchange factor activity
GO:0008270	1.20E-02	zinc ion binding
GO:0003676	1.27E-02	nucleic acid binding
GO:0003730	1.64E-02	mRNA 3'-UTR binding
GO:0000166	1.68E-02	nucleotide binding
GO:0016251	2.52E-02	general RNA polymerase II transcription factor activity
GO:0016887	2.52E-02	ATPase activity
GO:0004725	2.97E-02	protein tyrosine phosphatase activity
GO:0005089	3.17E-02	Rho guanyl-nucleotide exchange factor activity
GO:0005524	3.65E-02	ATP binding
GO:0004674	3.67E-02	protein serine/threonine kinase activity
GO:0042623	3.67E-02	ATPase activity, coupled
GO:0005102	3.67E-02	receptor binding
GO:0004930	4.02E-02	G-protein coupled receptor activity
GO:0005516	4.16E-02	calmodulin binding
GO:0003713	4.98E-02	transcription coactivator activity

Reported P-values are corrected for a false-discovery rate of 0.05.

TABLE S6

Signficantly-enriched Gene Ontology categories in top 1% 1kb copy-number variable (CNV) regions.

Biological Process

GO accession	P-value	Description
GO:0006355	3.58E-10	regulation of transcription, DNA-dependent
GO:0007417	1.10E-08	central nervous system development
GO:0045449	6.87E-08	regulation of transcription
GO:0007507	3.73E-06	heart development
GO:0007155	1.98E-05	cell adhesion
GO:0008585	3.54E-05	female gonad development
GO:0006813	3.81E-05	potassium ion transport
GO:0048477	1.20E-04	oogenesis
GO:0001700	1.55E-04	embryonic development via the syncytial blastoderm
GO:0007399	5.02E-04	nervous system development
GO:0007304	5.70E-04	chorion-containing eggshell formation
GO:0007619	5.97E-04	courtship behavior
GO:0008587	7.14E-04	imaginal disc-derived wing margin morphogenesis
GO:0007411	7.35E-04	axon guidance
GO:0007186	8.06E-04	G-protein coupled receptor protein signaling pathway
GO:0007494	8.06E-04	midgut development
GO:0007480	9.94E-04	imaginal disc-derived leg morphogenesis
GO:0007165	1.03E-03	signal transduction
GO:0007476	1.03E-03	imaginal disc-derived wing morphogenesis
GO:0008360	1.03E-03	regulation of cell shape
GO:0045475	1.33E-03	locomotor rhythm
GO:0007422	1.39E-03	peripheral nervous system development
GO:0007224	1.53E-03	smoothened signaling pathway
GO:0048190	1.72E-03	wing disc dorsal/ventral pattern formation
GO:0042048	1.80E-03	olfactory behavior
GO:0002121	1.86E-03	inter-male aggressive behavior
GO:0007623	1.95E-03	circadian rhythm
GO:0008354	1.95E-03	germ cell migration
GO:0008104	2.26E-03	protein localization
GO:0007455	2.51E-03	eye-antennal disc morphogenesis
GO:0035023	2.87E-03	regulation of Rho protein signal transduction
GO:0016318	2.87E-03	ommatidial rotation
GO:0007400	2.87E-03	neuroblast fate determination
GO:0007015	2.96E-03	actin filament organization
GO:0008285	3.38E-03	negative regulation of cell proliferation
GO:0009987	5.31E-03	cellular process
GO:0006911	5.79E-03	phagocytosis, engulfment

GO:0016321	8.07E-03	female meiosis chromosome segregation
GO:0009790	8.83E-03	embryonic development
GO:0007419	8.83E-03	ventral cord development
GO:0048749	8.83E-03	compound eye development
GO:0007268	9.23E-03	synaptic transmission
GO:0007517	9.23E-03	muscle organ development
GO:0007163	9.23E-03	establishment or maintenance of cell polarity
GO:0008407	9.23E-03	bristle morphogenesis
GO:0016567	9.23E-03	protein ubiquitination
GO:0007498	1.17E-02	mesoderm development
GO:0006508	1.17E-02	proteolysis
GO:0006366	1.32E-02	transcription from RNA polymerase II promoter
GO:0016481	1.32E-02	negative regulation of transcription
GO:0007423	1.34E-02	sensory organ development
GO:0006350	1.42E-02	transcription
GO:0007420	1.42E-02	brain development
GO:0030707	1.71E-02	ovarian follicle cell development
GO:0007611	1.96E-02	learning or memory
GO:0030036	1.96E-02	actin cytoskeleton organization
GO:0006357	1.97E-02	regulation of transcription from RNA polymerase II promoter
GO:0016055	1.97E-02	Wnt receptor signaling pathway
GO:0007269	2.02E-02	neurotransmitter secretion
GO:0006470	2.03E-02	protein amino acid dephosphorylation
GO:0006468	2.05E-02	protein amino acid phosphorylation
GO:0007242	2.12E-02	intracellular signaling cascade
GO:0007219	2.20E-02	Notch signaling pathway
GO:0035071	2.24E-02	salivary gland cell autophagic cell death
GO:0007409	2.34E-02	axonogenesis
GO:0019730	2.34E-02	antimicrobial humoral response
GO:0000381	2.49E-02	regulation of alternative nuclear mRNA splicing, via spliceosome
GO:0007298	2.68E-02	border follicle cell migration
GO:0008355	3.04E-02	olfactory learning
GO:0000122	3.30E-02	negative regulation of transcription from RNA polymerase II promoter
GO:0016319	3.57E-02	mushroom body development
GO:0007017	4.20E-02	microtubule-based process
GO:0008283	4.71E-02	cell proliferation
GO:0006811	4.82E-02	ion transport

Molecular Function

GO accession	P-value	Description
GO:0003700	1.71E-13	transcription factor activity
GO:0005515	4.61E-13	protein binding
GO:0008270	2.69E-11	zinc ion binding
GO:0043565	7.85E-08	sequence-specific DNA binding
GO:0004879	3.69E-06	ligand-dependent nuclear receptor activity
GO:0003704	5.63E-06	specific RNA polymerase II transcription factor activity
GO:0005041	6.03E-06	low-density lipoprotein receptor activity
GO:0030528	8.75E-06	transcription regulator activity
GO:0003702	1.03E-04	RNA polymerase II transcription factor activity
GO:0008239	1.91E-04	dipeptidyl-peptidase activity
GO:0003779	1.91E-04	actin binding

GO:0003707	3.69E-04	steroid hormone receptor activity
GO:0000166	4.35E-04	nucleotide binding
GO:0016566	5.63E-04	specific transcriptional repressor activity
GO:0005249	5.95E-04	voltage-gated potassium channel activity
GO:0003729	6.33E-04	mRNA binding
GO:0004930	9.72E-04	G-protein coupled receptor activity
GO:0005089	3.13E-03	Rho guanyl-nucleotide exchange factor activity
GO:0005509	3.30E-03	calcium ion binding
GO:0004872	5.40E-03	receptor activity
GO:0004871	6.59E-03	signal transducer activity
GO:0005200	6.59E-03	structural constituent of cytoskeleton
GO:0004674	7.26E-03	protein serine/threonine kinase activity
GO:0004222	8.70E-03	metalloendopeptidase activity
GO:0042803	8.70E-03	protein homodimerization activity
GO:0008188	9.50E-03	neuropeptide receptor activity
GO:0046982	1.52E-02	protein heterodimerization activity
GO:0004725	2.36E-02	protein tyrosine phosphatase activity
GO:0005198	3.55E-02	structural molecule activity
GO:0004842	3.66E-02	ubiquitin-protein ligase activity
GO:0051082	3.66E-02	unfolded protein binding
GO:0016564	3.66E-02	transcription repressor activity
GO:0003676	4.08E-02	nucleic acid binding

Reported P-values are corrected for a false-discovery rate of 0.05.

TABLE S7

The highly-differentiated circadian-regulation gene *norpA* and its known genetic-interaction partners.

FB accession	Gene	Functional Description
FBgn0004625*	norpA	phospholypase C, temperature synchronization of circadian clock
FBgn0025680*	Cry	response to light, circadian clock synchronization
FBgn0003218*	RdgB	required for light response
FBgn0000121	Arr2	deactivation of rhodopsin-mediated signaling
FBgn0039774	CDase	synaptic transmission
FBgn0002524	lace	–
FBgn0002940*	ninaE	response to light
FBgn0085373*	rdgA	rhodopsin signaling pathway

Asterisks (*) indicate that the gene is in the top 2.5% of 1kb F_{ST} windows.

FILE S1

Corrections for pooled sampling

The nature of our experimental design creates additional noise that we must correct for in our population genetic estimates. In particular, the pooled DNA sequencing design of this manuscript creates a second level of binomial sampling, beyond what is associated with the "normal" population genetic survey. Throughout we assume that a sample of size n chromosomes is taken from nature and pooled for sequencing. This sequencing is performed to depth m which may be variable among loci/sites. Conditioning on a population frequency of the A_1 allele p, the probability of sampling i out of n A_1 alleles in our initial sample is simply binomial with parameters n and p. Thus the expected value of the sample frequency $E(i/n) = p$ and the second moment is $E((i/n)^2) = \frac{p(1-p)}{n} + p^2$.

We will first derive similar results for the pooled sampling design, and then move on to estimation of population genetic statistics. The probability of sampling k A_1 alleles out of m in our pooled sequences conditional upon having sampled i out of n in our initial sampling is again binomial, this time with parameters m and i/n. Thus the probability of sequencing k out of m reads of the A_1 allele conditional upon the population allele frequency is

$$Prob(X = k|m, n, p) = \sum_{i=0}^{n} \binom{m}{k} (i/n)^k (1 - i/n)^{m-k} \binom{n}{i} p^i (1-p)^{n-i}. \quad (1)$$

The expected value of the frequency of the allele in our sequenced sample, k/m, can then be easily found through the use of conditional expectations

$$\begin{aligned} E\{k/m\} &= E\{E\{k/m|i/n\}\} \\ &= \sum_{i=0}^{n} E\{k/m|i/n\} \times Prob(i) \\ &= \sum_{i=0}^{n} \frac{m(i/n)}{m} \binom{n}{i} p^i (1-p)^{n-i} = p \end{aligned} \quad (2)$$

We can find the second moment through similar means

$$\begin{aligned} E\{(k/m)^2\} &= E\{E\{(k/m)^2|i/n\}\} \\ &= \sum_{i=0}^{n} E\{(k/m)^2|i/n\} \times Prob(i) \\ &= \sum_{i=0}^{n} \left\{ \frac{(i/n)(1-i/n)}{m} + (i/n)^2 \right\} \times Prob(i) \\ &= \sum_{i=0}^{n} \frac{(i/n)(1-i/n)}{m} \times Prob(i) + \sum_{i=0}^{n} (i/n)^2 \times Prob(i) \end{aligned} \quad (3)$$

There are two terms in equation 3. This second term is immediately recognizable as the second moment that we examined above (i.e. $\frac{p(1-p)}{n} + p^2$). The first term after a bit of algebra turns into

$$\frac{1}{m}\{p(1-p) - \frac{p(1-p)}{n})\}$$

putting it all together, the expectation of the second moment, conditional upon the population allele frequency is

$$E\{(k/m)^2\} = p(1-p)/m - p(1-p)/mn + p(1-p)/n$$

With that result in hand we are now ready to write down the expectation of heterozygosity ($H = 2p(1-p)$) given our population allele frequency

$$\begin{aligned} E\{H\} &= E\{2p(1-p)\} = 2(E\{p\} - E\{p^2\}) \\ &= 2\Big(p - p(1-p)/m + p(1-p)/mn - p(1-p)/n\Big) \\ &= 2p(1-p)((n-1)/n)((m-1)/m) \end{aligned} \quad (4)$$

This leads to a simple bias correction on our estimates of heterozygosity which is $n/(n-1) \times m/(m-1)$. Figure S7 shows coalescent simulation results, where we generated samples from the standard coalescent model, and then resampled chromosomes with replacement to a coverage depth m. We then applied both the "double" correction derived here and the standard single correction. As can be seen in that figure, we do indeed have an unbiased estimator of heterozygosity if we correct for both the original size of our pooled sample and the coverage.

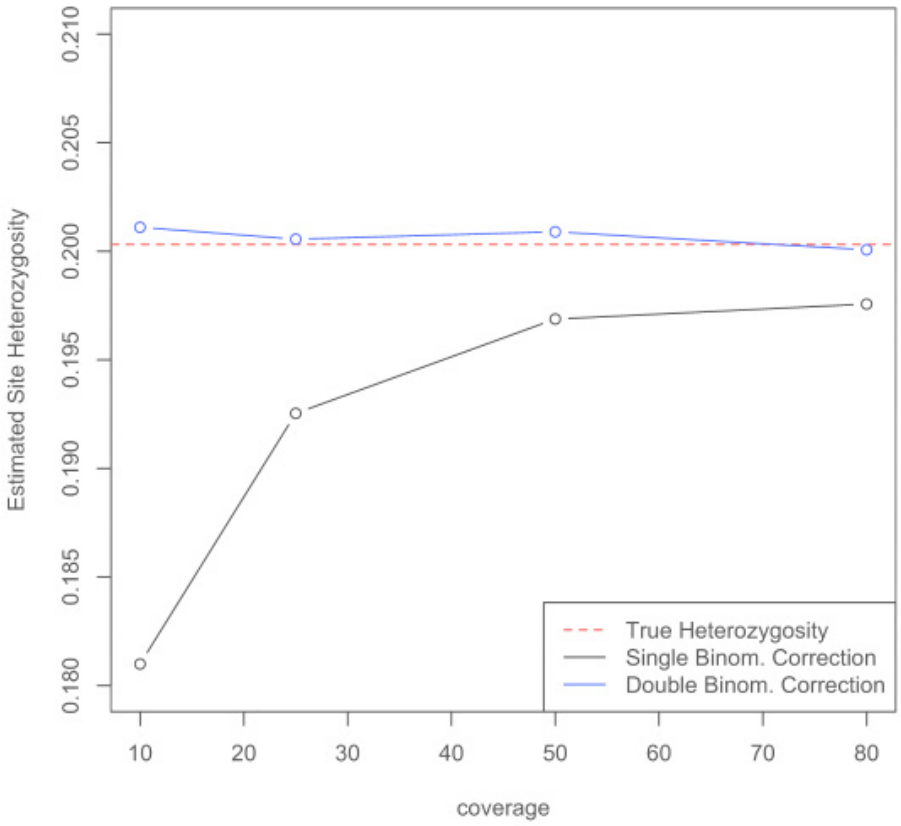

FIGURE S5.—Simulation results showing the corrected heterozygosity (eqn 4) is effective across a range of coverages used in this manuscript

Estimating θ

Of interest to us is coming up with an unbiased estimator of $\theta = 4Nu$ in the face of our pooled sampling strategy. Recently (Futschik and Schlotterer 2010) have done quite a bit of work on this problem, and they were able to come up with corrected estimators for $\theta_p i$ and θ_w. Here we derive ostensibly similar results through different means, and arrive at a generalized correction for pooled sampling which allows for construction of arbitrary estimators of θ as linear combinations of the site frequency spectrum using the system of (Achaz 2008).

We start by generalizing equation 1 of the supplement across the sample site frequency spectrum (SFS) expected under the standard neutral model. The probability of observing an allele segregating at frequency i out of n in a standard sample is $Prob(i|n) = 1/ia_n$, where $a_n = \sum_{i=1}^{n-1} 1/i$ (Ewens 2004). Thus the probability of observing an allele at frequency k out of m reads in our pooled sequence sample unconditional on the population

frequency is

$$\begin{aligned} Prob(k|m,n) &= \sum_{i=1}^{n-1} Prob(K|m,n,i)Prob(i) \\ &= \sum_{i=1}^{n-1} \binom{m}{k} (i/n)^k (1-i/n)^{m-k} (1/ia_n) \end{aligned} \tag{5}$$

This expression allows us to write down the expected number of sites segregating at frequency k out of m, call it Y_k as a function of the mutation rate θ. Conditioning on the expected number of segregating sites S in our initial sample of size n allows us to write down the expected values of the Y_ks as

$$\begin{aligned} E\{Y_k\} &= E\{S\} \times Prob(k|m,n) \\ &= \theta a_n \sum_{i=1}^{n-1} \binom{m}{k} (i/n)^k (1-i/n)^{m-k} (1/ia_n) \end{aligned} \tag{6}$$

To check the accuracy of this expression we performed coalescent simulations with a specified sample size n and mutation rate θ. The initial site frequency spectrum was recorded and then transformed to mimic our pooled sequencing by sampling alleles at each segregating site with replacement. This yields a transformed SFS Y – see figure S6.

Given the accuracy of our correction for the SFS we move on to derive a bias corrected estimation scheme for θ. In particular we note that rearrangement of equation 6 suggests a moment estimator of the type derived in Fu (1995),

$$\hat{\theta} = \frac{Y_i}{a_n} \frac{1}{Prob(k|m,n)}$$

Achaz (2009) noted that linear combinations of the SFS can be used to derive new estimators of θ given some arbitrary weighting scheme. In this context we can write down the bias corrected version of Achaz's generic estimator as

$$\hat{\theta}_\omega = \frac{1}{a_n \sum_k \omega_k} \sum_{k=1}^{m-1} \omega_k Y_k \frac{1}{Prob(k|m,n)}. \tag{7}$$

In this manuscript we focus attention on Tajima's nucleotide diversity ($\hat{\theta}_\pi$) and Fay and Wu's ($\hat{\theta}_H$) (Tajima 1983; Fay and Wu 2000). To show the potential generality of our correction scheme we present coalescent simulation results as before where we have estimated θ using six different weighting schemes: $\omega_{\pi,i}$, $\omega_{H,i}$, and $\omega_{W,i}$, each with and without use of singletons

$$\begin{aligned}\omega_{\pi,i} &= n - i \\ \omega_{H,i} &= i \\ \omega_{W,i} &= i^{-1}\end{aligned}$$

In the case where singletons are ignored each of the $\omega_1 = 0$. As can be seen in Figure S7 our bias corrected estimates are quite accurate, thus the framework we have introduced here should be general.

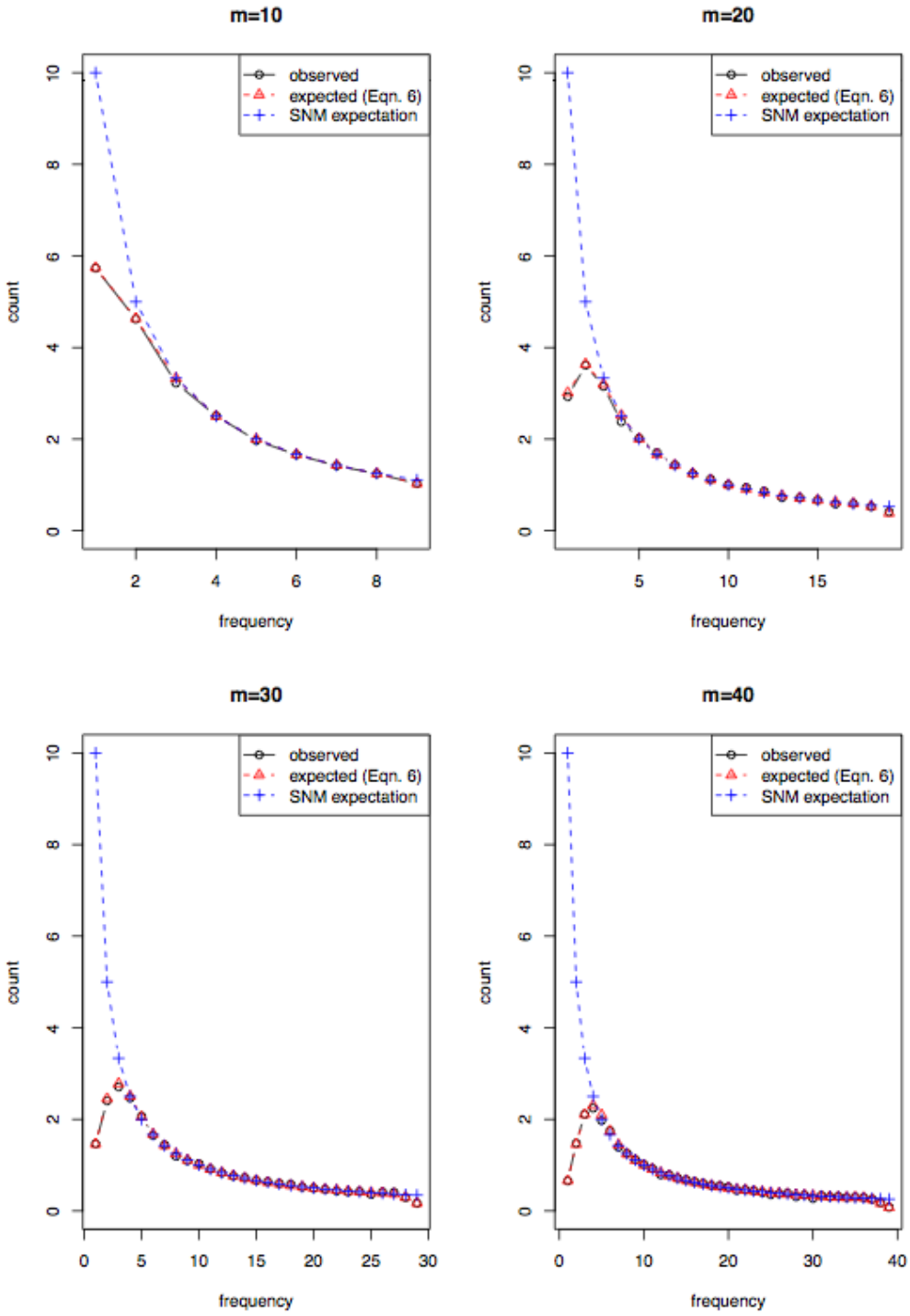

FIGURE S6.—Simulation results showing the correspondence between the observed and expected site frequency spectrum as m the sequencing depth changes. 1000 coalescent simulations were run with n=10 and θ=10. The expected values in red are derived from equation 6. Shown for comparison in blue are the expectations under the standard neutral model.

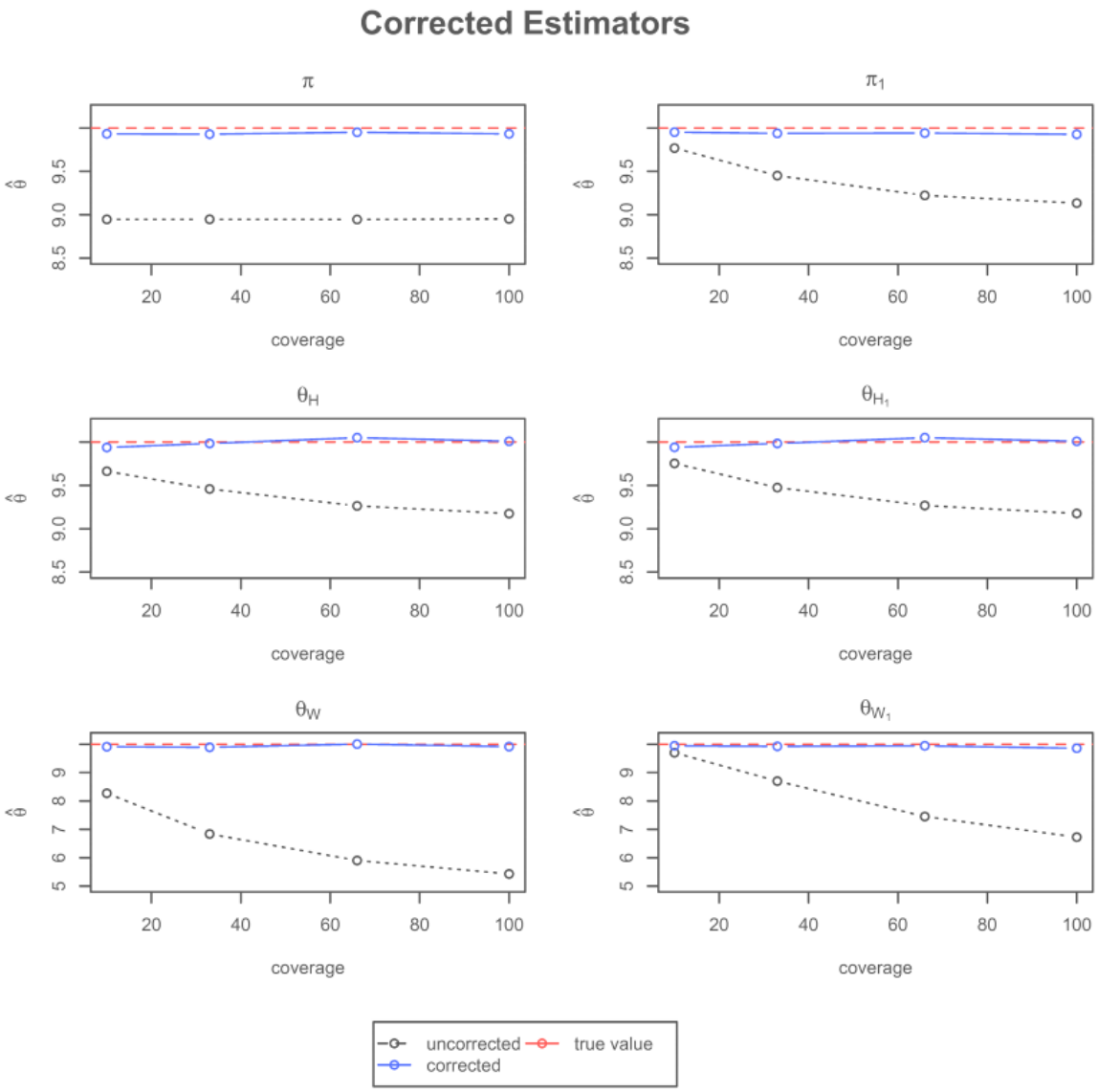

FIGURE S7.—Simulation results showing the performance of our bias corrected estimators of θ. 1000 coalescent simulations were run with $n = 40$ and $\theta = 10$. Uncorrected estimates are shown in black.

REFERENCES

ACHAZ, G., 2008, Jul) Testing for neutrality in samples with sequencing errors. Genetics *179* (3): 1409–1424.

ACHAZ, G., 2009 Frequency spectrum neutrality tests: one for all and all for one. Genetics *183* (1): 249–58.

EWENS, W. J., 2004 *Mathematical population genetics* (2nd ed ed.), Volume v. 27. New York: Springer.

FAY, J. C. and C. I. WU, 2000, Jul) Hitchhiking under positive Darwinian selection. Genetics *155* (3): 1405–13.

FU, Y. X., 1995 Statistical Properties of Segregating Sites. Theoretical Population Biology **48**: 172–197.

FUTSCHIK, A. and C. SCHLOTTERER, 2010 Massively Parallel Sequencing of Pooled DNA Samples–The Next Generation of Molecular Markers. Genetics.

TAJIMA, F., 1983, Oct) Evolutionary relationship of DNA sequences in finite populations. Genetics *105* (2): 437–60.