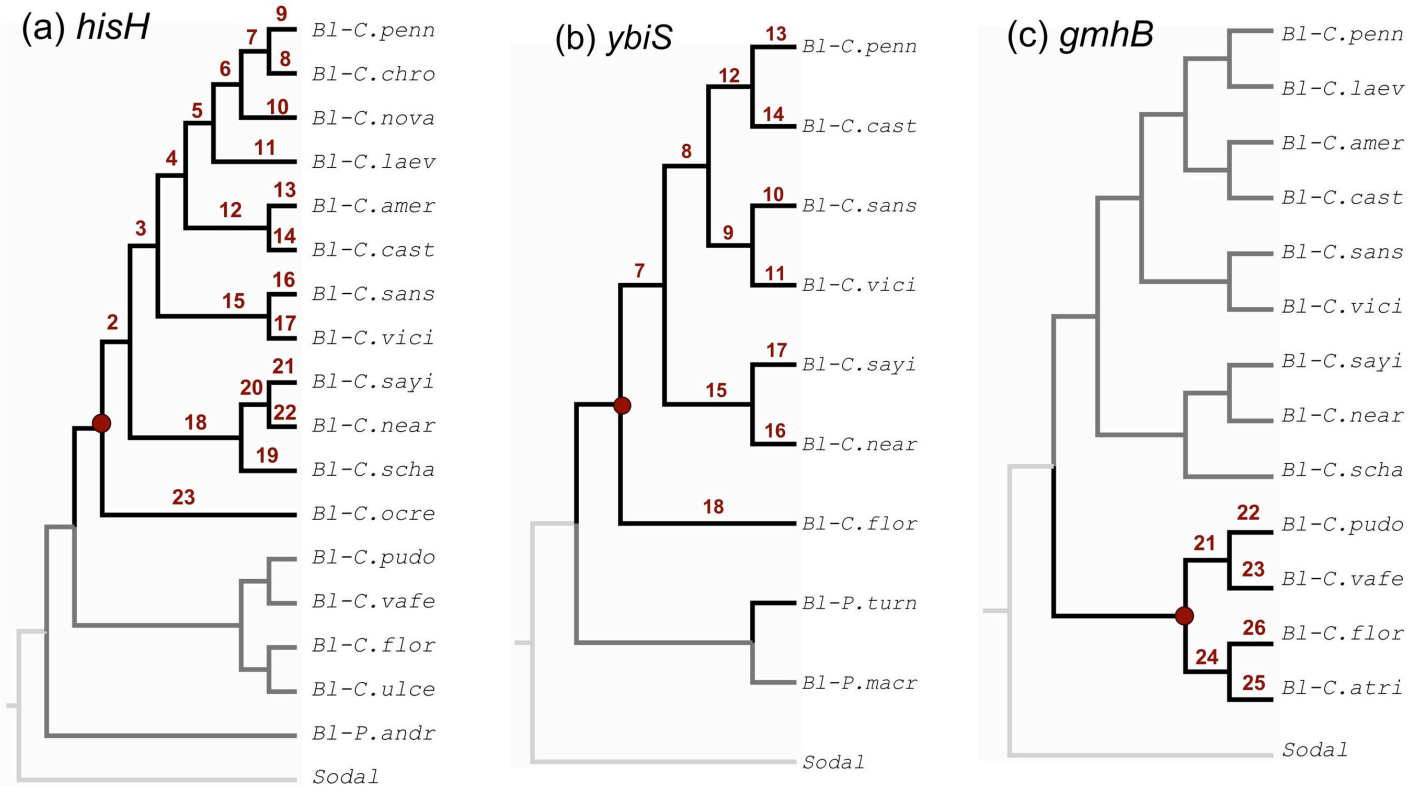
Supplementary Figure 1



Supplementary Figure 1.

Branches across which four-fold degenerate substitutions were tracked, in order to test whether homopolymeric tracts show unusual nucleotide conservation. The branches considered in this analysis are labeled with red ID numbers (reflecting the ID's in Supplementary Table 1). These branches span the clade that retains a frameshift within the homopolymeric region, marked with a red circle. In total, we included 22 branches for *hisH*, 11 branches for *ybiS*, and 6 branches for *gmhB*. In the cladograms presented, branch lengths do not reflect sequence divergence levels.

Supplementary Table 1.
Chance that no substitutions occur along homopolymeric tracts of frameshifted clades, based on observed rates of four-fold degenerate changes across branches.

hisH

Branch ID[a]	$N_{4\text{-fold}}$[b]	$A_{4\text{-fold}}$[b]	$A_{4\text{-fold}}$ -> N along branch	$A_{4\text{-fold}}$ -> G	$A_{4\text{-fold}}$ -> C	$A_{4\text{-fold}}$ -> T	($A_{4\text{-fold}}$ -> N) /$A_{4\text{-fold}}$	Prob no change at an $A_{4\text{-fold}}$ site[c]	Prob no change at 11 $A_{4\text{-fold}}$ sites[d]	($A_{4\text{-fold}}$ -> G) /$A_{4\text{-fold}}$	Prob no ->G change at a $A_{4\text{-fold}}$ site[c]	Prob no ->G changes at 3 $A_{4\text{-fold}}$ sites[d]
2	58	29	2		1	1	0.069	0.931	0.456	0.000		
3	59	29	1			1	0.034	0.966	0.680	0.000		
4	60	29	1	1			0.034	0.966	0.680	0.034	0.966	0.900
5	61	28	2	2			0.071	0.929	0.443	0.071	0.929	0.801
6	64	28	0	0			0.000			0.000		
7	64	28	3	2		1	0.107	0.893	0.287	0.071	0.929	0.801
8	64	27	1			1	0.037	0.963	0.660	0.000		
9	64	27	0				0.000					
10	64	28	5	3	1	1	0.179	0.821	0.115	0.107	0.893	0.712
11	64	28	1			1	0.036	0.964	0.670	0.000		
12	61	28	5	5			0.179	0.821	0.115	0.179	0.821	0.554
13	61	24	1	1			0.042	0.958	0.626	0.042	0.958	0.880
14	61	24	0				0.000					
15	60	29	6	2	2	2	0.207	0.793	0.078	0.069	0.931	0.807
16	62	23	3	2	1		0.130	0.870	0.215	0.087	0.913	0.761
17	62	23	4	3		1	0.174	0.826	0.122	0.130	0.870	0.658
18	59	29	0				0.000					
19	59	29	3	1		2	0.103	0.897	0.301	0.034	0.966	0.900
20	59	29	1			1	0.034	0.966	0.680	0.000		
21	55	27	6	2	2	2	0.222	0.778	0.063	0.074	0.926	0.794
22	55	27	4	2	1	1	0.148	0.852	0.171	0.074	0.926	0.794
23	58	29	7	1	3	3	0.241	0.759	0.048	0.034	0.966	0.900
Across all branches[e]:									**2.127E-11**			**4.131E-02**

ybiS

7	102	45	2	1		1	0.044	0.956	0.664	0.022	0.978	0.935
8	102	44	0				0.000			0.000		
9	102	41	3	2		1	0.073	0.927	0.505	0.049	0.951	0.861
10	102	39	5	4		1	0.128	0.872	0.291	0.103	0.897	0.723
11	102	39	7	5	2		0.179	0.821	0.169	0.128	0.872	0.663
12	102	41	4	4			0.098	0.902	0.397	0.098	0.902	0.735
13	102	39	12	6	4	2	0.308	0.692	0.037	0.154	0.846	0.606
14	102	39	9	3	2	4	0.231	0.769	0.094	0.077	0.923	0.787
15	102	44	6	2		4	0.136	0.864	0.267	0.045	0.955	0.870
16	107	47	8	7		1	0.170	0.830	0.187	0.149	0.851	0.616
17	107	47	7	4	2	1	0.149	0.851	0.234	0.085	0.915	0.766
18	102	45	13	1	1	11	0.289	0.711	0.046	0.022	0.978	0.935
Across all branches[e]:									**1.220E-08**			**5.179E-02**

gmhB

Branch ID[a]	$N_{4\text{-fold}}$[b]	$T_{4\text{-fold}}$[b]	#$T_{4\text{-fold}}$ -> N	#$T_{4\text{-fold}}$ -> G	#$T_{4\text{-fold}}$ -> C	#$T_{4\text{-fold}}$ -> A	($T_{4\text{-fold}}$ -> N) / $T_{4\text{-fold}}$	Prob no change at a $T_{4\text{-fold}}$ site[c]	Prob no change at 9 $T_{4\text{-fold}}$ sites[d]	($T_{4\text{-fold}}$ -> C) / $T_{4\text{-fold}}$	Prob no ->C change at a $T_{4\text{-fold}}$ site[c]	Prob no ->C changes at 3 $T_{4\text{-fold}}$ sites[d]
21	52	21	0				0.000			0.000		
22	50	24	4	1		3	0.167	0.833	0.194	0.000		
23	50	24	9	2	5	2	0.375	0.625	0.015	0.208	0.792	0.496
24	52	21	5	1	2	2	0.238	0.762	0.087	0.095	0.905	0.741
25	52	25	0				0.000			0.000		
26	52	25	0				0.000			0.000		
Across all branches[e]:									**2.440E-04**			**3.675E-01**

[a]Branch IDs are labeled along trees in Supplementary Figure 1.

[b]Values refer to the number of four-fold degenerate sites (N_4-fold) and number of four-fold degenerate A's (or T's) in the reconstructed ancestor for that branch.

[c]The chance of no substitution at given site was calculated as one minus the frequency of that change (column to the left). These values were not calculated when the relevant change did not occur along a branch, as this would have suggested an 100% probability of no change.

[d]The chance of no substitutions at a particular number of sites (the tract length, and number of synonymous sites in the tract) was calculated as thechance of no substitution at one site (column to the left) raised to the power of the site number considered.

[e]The chance of no substitutions across all branches was calculated as the product of the above branch-specific values.

Supplementary Table 2. Maximum likelihood estimates of nonsynonymous/synonymous divergences, ω, for regions (a) upstream or (b) downstream of the homopolymeric tract. Branch models (presented in Table 3 of the main text) were also performed across portions of the genes upstream and downstream of the homopolymeric tract. In PAML, ω was calculated as a single ratio across all *Blochmannia* branches in the phylogeny, using the M0 (one-ratio) model, and a two-ratio model allowing distinct ω values for branches with and without the frameshift. A likelihood ratio test was used to test whether lnL of the two models differed significantly. With the exception of the upstream region of *ybiS*, ω did not differ significantly between frameshifted and intact clades. The *ybiS* result may be unreliable due to high dS (>3) along the single lineage lacking the frameshift.

(b) Upstream of the tract region

Gene	bp used	One-ratio model		Two-ratio model			Likelihood Ratio Test	
		ω	lnL	ω (frameshift)	ω (intact)	lnL	2ΔL	
hisH	168	0.116	-1292.946	0.125	0.099	-1292.710	0.473	n.s.
gmhB	75	0.068	-461.587	0.030	0.093	-460.176	2.822	n.s.
ybiS	414	0.134	-2582.626	0.034	0.151	-2578.752	7.749	p<0.01

(b) Downstream of the tract region

Gene	bp used	One-ratio model		Two-ratio model			Likelihood Ratio Test	
		ω	lnL	ω (frameshift)	ω (intact)	lnL	2ΔL	
hisH	399	0.105	-3309.565	0.096	0.122	-3308.943	1.245	n.s.
gmhB	456	0.126	-3054.404	0.130	0.124	-3054.377	0.053	n.s.
ybiS	468	0.086	-2957.424	0.089	0.035	-2956.426	1.995	n.s.

*At *ybiS*, the lineage leading to the intact *P. tuneri* sequence had a lower ω value; however, the very high dS (>3) along this lineage confounds the comparison (see text).