

# Long-term microclimatic stress causes rapid adaptive radiation of *kaiABC* clock gene family in a cyanobacterium, *Nostoc linckia*, from “Evolution Canyons” I and II, Israel

Volodymyr Dvornyk<sup>\*†</sup>, Oxana Vinogradova<sup>‡</sup>, and Eviatar Nevo<sup>\*§</sup>

<sup>\*</sup>Institute of Evolution, University of Haifa, Mount Carmel, 31905 Haifa, Israel; and <sup>†</sup>N.G. Kholodny Institute of Botany, National Academy of Sciences of Ukraine, vul. Tereshchenkivska 2, Kiev, Ukraine

Contributed by Eviatar Nevo, December 23, 2001

Cyanobacteria are the only prokaryotes known thus far possessing regulation of physiological functions with approximate daily periodicity, or circadian rhythms, that are controlled by a cluster of three genes, *kaiA*, *kaiB*, and *kaiC*. Here we demonstrate considerably higher genetic polymorphism and extremely rapid evolution of the *kaiABC* gene family in a filamentous cyanobacterium, *Nostoc linckia*, permanently exposed to the acute natural environmental stress in the two microsite evolutionary models known as “Evolution Canyons,” I (Mount Carmel) and II (Upper Galilee) in Israel. The family consists of five distinct subfamilies (*kaiI–kaiV*) comprising at least 20 functional genes and pseudogenes. The obtained data suggest that the duplications of *kai* genes have adaptive significance, and some of them are evolutionarily quite recent ( $\approx 80,000$  years ago). The observed patterns of within- and between-subfamily polymorphisms indicate that positive diversifying, balancing, and purifying selections are the principal driving forces of the *kai* gene family's evolution.

The problem of the relative importance of various driving forces in genome evolution requires a comprehensive critical testable approach. The challenging question is that of how selective or nonselective factors influence the evolution of organisms living under contrasting ecological conditions.

Lower Nahal Oren, Mount Carmel, and Lower Nahal Keziv, western Upper Galilee, Israel, known as “Evolution Canyons” (ECs) I and II, respectively (1–3), represent microsites with highly contrasting environmental conditions on opposite slopes separated by only 300–400 m at the top and 50–100 m at the bottom. They were eroded in the Plio-Pleistocene period. The “African” south-facing slopes (SFSs) receive up to 600% more solar radiation and are warmer, drier, and spatiotemporally more heterogeneous and fluctuating than “European” north-facing slopes (NFSs). Such ecological contrasts on a microscale provide an excellent opportunity to test various hypotheses regarding the influence of natural environmental stress on patterns of genome evolution and shaping genetic variability of populations (1), adaptive radiation, and incipient speciation across life (2, 3).

Cyanobacteria (blue-green algae, *Cyanoprokaryota*, *Cyanophyta*) can occupy various ecotopes and have high ecological adaptability (4) and thus are suitable model organisms for testing various predictions about probable evolutionary forces affecting patterns of genetic polymorphism under contrasting ecological conditions. Thus far, extensive studies have been done to ascertain molecular and genetic mechanisms of cyanobacterial responses to various stresses (see refs. 5–9).

However, despite the extensive studies of cyanobacteria, data about the effect of long-time natural environmental stress on the level and patterns of nucleotide polymorphism in these prokaryotes are lacking. Such data can highlight the evolutionary forces contributing to adaptive genetic radiation and speciation. This issue is of particular interest, because it pertains to processes shaping genetic variation at intra- and interpopulation levels under sharp

environmental gradients at microscales, where microevolution can be tracked effectively (3).

According to the niche-width variation hypothesis (10) and the environmental theory of genetic diversity (11–13), populations from spatiotemporally more variable and/or ecologically stressful habitats should display higher morphological and genetic polymorphism. Filamentous cyanobacterium *Nostoc linckia* is a ubiquitous species inhabiting both slopes of the two ECs (14). Because of its sessile character, the cyanobacterium is exposed permanently to fluctuating complex environmental stress (light, temperature, desiccation, etc.) and thus is expected to follow the above predictions. Our recent study showed significantly higher levels of intra- and interpopulation amplified fragment length (15) and HIP1 (16) polymorphisms of *Nostoc* on the SFS compared with the NFS.

On the other hand, according to the neutral theory, nucleotide polymorphism in a population is determined by effective population size and the mutation rate,  $\theta = 4N_e\mu$  (17). It suggests that a large effective population size and high mutation rate should increase the level of nucleotide variation. In the ECs, the much higher level of UV radiation received by the SFS may result in a higher mutation rate and, along with the abundance of the cyanobacterium, a higher nucleotide polymorphism when compared with the NFS.

Among environmental factors affecting biota at the ECs, the level of solar and UV radiation seems to be the most stressful and contrasting for the slopes. We therefore hypothesized that genes related to light-dependent physiological and biochemical processes should be suitable to test the above hypotheses and thus chose genes regulating circadian rhythms in *N. linckia*.

Regulation of physiological functions with approximate daily periodicity, or circadian rhythms, is a common feature of eukaryotes. Among prokaryotes, only cyanobacteria show this rhythmicity, which is regulated by a cluster of three genes, *kaiA*, *kaiB*, and *kaiC* (18, 19). The *kai* genes have been studied most extensively in the unicellular cyanobacterium *Synechococcus*. They were shown to operate as a single unit and to have a complicated mode of regulation; *kaiA* positively regulates *kaiBC* promoter, whereas overexpression of *kaiC* represses it (18, 20). Recently *kaiC* was reported to have two *kaiA*-binding domains (21) and to be a crucial component of clock precession in cyanobacteria (22). Circadian pacemakers in cyanobacteria were shown to influence the expression of a wide variety of genes including those regulating nitrogen fixation (23), cell division (24), and other metabolic processes (25,

Abbreviations: EC, Evolution Canyon; SFS, south-facing slope; NFS, north-facing slope; CBI, codon bias index.

<sup>†</sup>Present address: Osteoporosis Research Center, Creighton University, 601 North 30th Street, Suite 6730, Omaha, NE 68131. E-mail: dvornyk@creighton.edu.

<sup>§</sup>To whom reprint requests should be addressed. E-mail: nevo@research.haifa.ac.il.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

26). Furthermore, close matching between the endogenous clock and the temporal environmental cycle improves the reproductive fitness of cyanobacteria (27). In all cyanobacteria studied thus far, the *kaiABC* cluster was reported as a single copy (18, 19).

In the present study we address several problems. First, what are the level and patterns of sequence polymorphism in *kai* genes of *N. linckia* from the ECs microsites? Second, what evolutionary factors contribute most significantly to the observed level of the polymorphism? Third, is there any adaptive significance to the observed polymorphism?

## Materials and Methods

**Cyanobacterial Strains.** Five cyanobacterial strains from each of 12 sampling stations (three stations on each slope of both ECs) were analyzed. The strains were maintained on BG-11-0 media (Sigma) with 1.5% agar. For DNA isolation, the cultures were grown on 50 ml of BG-11-0 liquid media for 1 month to be prepared for the subsequent DNA isolation.

**DNA Isolation, Amplification, and Sequencing.** Total genomic DNA from the cyanobacterial lines was isolated by the miniprep procedure using CTAB (28) with some minor modifications. PCR primers were designed based on the published partial sequences of the *kaiBC* region in *Cylindrospermum* PCC 7417 (GenBank accession no. AF222605) and *Nodularia* PCC 73104 (AF222602). The amplifications were performed with forward primer kF37 (5'-AGC CGA AGA AGA TAA AAT-3') and reverse primer kR910 (5'-GAC GTT CTC CTT CTA AAA-3'). The names of the primers refer to the respective nucleotide position in the *Nodularia* PCC 73104 *kaiBC* sequence. PCR was made in a total of 20  $\mu$ l of reaction mix containing 1 unit of REDTaq DNA polymerase (Sigma), 10 ng of template, 1 $\times$  REDTaq PCR buffer, 1.5 mM MgCl<sub>2</sub>, 250  $\mu$ M dNTPs, and 0.25  $\mu$ M of each primer. The PCR program consisted of one cycle at 94°C for 3 min followed by 45 cycles of 1 min at 94°C, 50 sec at 50°C, and 50 sec at 72°C and by final extension at 72°C for 10 min. The PCR product then was purified from 1% agarose gel with the QIAquick gel-extraction kit (Qiagen) and either used for direct sequencing (BigDye sequencing kit, Perkin-Elmer Applied Biosystems) or cloned in pGEM-T Easy vector (Promega). The cloned fragments were then isolated with QIAprep Spin miniprep kit (Qiagen) and sequenced with M13 primers. The fragments, differing by only a few singletons, were sequenced repeatedly after each additional PCR to avoid confusing with PCR errors. The newly determined *kaiBC* sequences of *N. linckia* were deposited in the GenBank, DDBJ, and EMBL databases (Table 1).

**Data Analyses.** The sequences were aligned with CLUSTALW (29). Synonymous ( $K_s$ ) and nonsynonymous ( $K_a$ ) nucleotide substitutions were calculated by using the modified Nei-Gajdoori method (30) with Jukes-Cantor correction and the transition/transversions ratio equal to 3. We also estimated a codon usage bias measured by the codon bias index (CBI; ref. 31). It may range from 0 (no codon bias) to 1 (maximum codon bias). Respective computations were performed by using DNASP, version 3.53 (32). The Poisson random field model with maximum likelihood estimation (33) was applied to test whether selection operates on shaping the observed nucleotide polymorphism. Phylogenetic inferences for the obtained genes were made by using the maximum likelihood approach (34) realized in DAMBE, version 4.0.57 (35). Published *kai* sequences of *Nodularia* PCC 73104 (AF222602), *Cylindrospermum* PCC 7417 (AF222605), and *Nostoc punctiforme* (preliminary sequence data were obtained from the Department of Energy Joint Genome Institute at [www.jgi.doe.gov/JGI\\_microbial/html/index.html](http://www.jgi.doe.gov/JGI_microbial/html/index.html)) were used as an outgroup. The statistical significance of the constructed trees was evaluated by bootstrapping with 1,000 replications.

**Table 1. List of the *kai* gene family members obtained from the analyzed cyanobacterial strains in the ECs**

Location	Station	Haplo-type	Allele	Total length (intergenic spacer) of the sequenced fragment, bp	GenBank accession no.
Lower Nahal	SFS1	A	<i>kaila</i> *	1039 (274)	AY051296
Oren (EC I)	D		<i>kaiIIId1</i>	1264 (499)	AY051303
			<i>kaiIIId2</i>	1264 (499)	AY051304
			<i>kaiIIId1</i>	871 (106)	AY051308
			<i>kaiIIId2</i>	878 (113)	AY051309
	E		<i>kaiIle</i>	1208 (443)	AY051301
			<i>kaiIe</i>	1039 (274)	AY051297
			$\psi$ <i>kaiIe</i>	1038 (274)	AY051298
			<i>kaiIIIe</i>	878 (113)	AY051310
	F		<i>kaiIIIff</i>	1264 (499)	AY051302
			<i>kaiIa</i>	1039 (274)	AY051296
			<i>kaiIIIff1</i>	871 (106)	AY051306
			<i>kaiIIIff2</i>	871 (106)	AY051307
Lower Nahal Keziv (EC II)	SFS2	C	<i>kaiIVb</i>	841 (76)	AY051312
			<i>kaiIVc</i>	841 (76)	AY051313
	SFS1	G	<i>kaiVg</i>	843 (78)	AY051317
			<i>kaiIIIg</i>	848 (83)	AY051311
			$\psi$ <i>kaiIlg</i>	1037 (274)	AY051300
			$\psi$ <i>kaiIIIg</i>	1228 (464)	AY051305
	SFS2	H	<i>kaiIVh</i>	843 (78)	AY051316
			<i>kaiIh</i>	1039 (274)	AY051299

\*Ubiquitous in all the locations unless specified, also occurs in haplotype F among the other *kai* gene family members. Roman numbers indicate gene subfamilies.

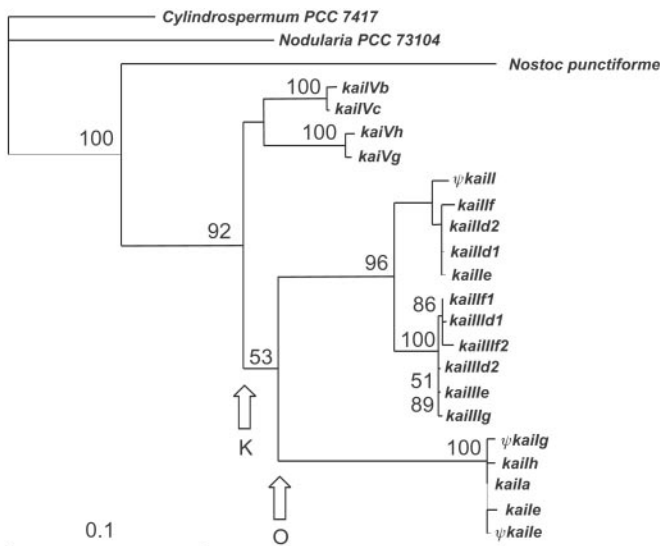
## Results

The amplified segment of the *kaiABC* cluster in *N. linckia* included parts of the *kaiB* and *kaiC* genes and an intergenic spacer,  $\approx$ 1 kb length in total. In the 60 strains analyzed (30 from each EC), we found at least eight haplotypes, which differed by either nucleotide sequence or the number of gene copies presented. In some strains from the most stressful upper SFS1 and SFS2 stations of both canyons, the *kai* cluster was represented by multiple copies, at least two or four (Table 1). The strains from all other sampling stations of both canyons had the same single completely monomorphic copy of the cluster, *kaiIa*. Multi-banded patterns of the amplicons observed on agarose gel referred to the indel polymorphism in the *kaiBC* intergenic spacer.

### Phylogeny of *kai* Multigene Family and the Time of the Duplications.

The constructed tree clearly indicates that the *kai* gene family is monophyletic and results from multiple duplication events (Fig. 1). All members of the family are subdivided into five groups, or subfamilies, by their nucleotide homology. Some cyanobacterial haplotypes may vary in the number of genes of different subfamilies (Table 1). The topology of the obtained phylogenetic tree suggests different times for the first duplication events occurring in the ECs (Fig. 1). A few duplications (for example, between *kaiIIId1* and *kaiIIId2*) were quite recent.

The molecular clock hypothesis was tested by the maximum likelihood test and was rejected if the genes of all the species (i.e., including those from the closely related *Nostocaceae*) were included in the test because of the faster rate of substitution in *N. linckia* *kai* gene family. However, the hypothesis was confirmed if this family was excluded. Having paleontological data about the origin of family *Nostocaceae* ( $\approx$ 1.6 billion years ago; refs. 36 and 37) and the average pairwise number of synonymous substitutions per synonymous site (30) between *Nodularia*, *Cylindrospermum*, and *N.*



**Fig. 1.** Maximum likelihood phylogenetic tree of the *kai* multigene family. The first duplication events in EC I and II are depicted as O and K, respectively. Bootstrap values (1,000 replications) are indicated at nodes. The bootstrap values <50% are not shown.

*punctiforme* *kai* genes ( $K_s = 0.706$ ), we thus could determine the neutral substitution rate in non-*linckian* lineages. It appeared to be  $2.21 \times 10^{-10}$  substitutions per site per year. The approximate time of the duplications was computed based on the data of geological history of the canyons (1, 38) and pairwise  $K_s$  values of the respective gene descendants. For example, having the first duplication event in Nahal Oren (Fig. 1 O)  $\approx 5$  million years ago and  $K_s$  between *kaiIe* and *kaiIIIe* equal to 0.420, we obtain the rate of  $4.20 \times 10^{-8}$  substitutions per site per year or  $\approx 190$  times higher than in the non-*linckian* lineages. Furthermore, taking  $K_s$  between *kaiIIId1* and *kaiIIId2* as 0.007, we finally can set the time of duplication to be  $\approx 80,000$  years ago.

**Nucleotide Polymorphism and Diversity Among Members of the *kai* Multigene Family.** Translated amino acid sequences of the fragments showed high similarity to the respective cyanobacterial sequences reported previously (19). Some of the obtained nucleotide sequences differed either by a few synonymous substitutions or by silent substitutions in intergenic spacers, and therefore their amino acid sequences were identical. A number of the *kaiBC* sequences proved to be pseudogenes caused by frameshift mutations. For example,  $\psi$ *kaiIlg* resulted from the deletion of a dinucleotide at position 75 of the *kaiB* gene,  $\psi$ *kaiIIlg* (from the deletion of a nucleotide at position 635 of *kaiC*).

**Table 2. Data on synonymous and nonsynonymous substitutions and nucleotide diversity in the different *kai* domains within the subfamilies of the *kai* gene family in *N. linckia***

<i>Kai</i> subfamily	No. of the member genes determined	<i>kaiB</i> (135 bp)			Intergenic spacer (79–502 bp)	<i>kaiC</i> (627 bp)			Total in coding regions		
		$K_s$	$K_a$	$\pi$ , SE	$\pi$	$K_s$	$K_a$	$\pi$ , SE	$K_s$	$K_a$	$\pi$ , SE
I	5	0.0131	0.0041	0.0061 $\pm$ 0.0023	0.0029	0.0192	0.0042	0.0077 $\pm$ 0.0017	0.0181	0.0042	0.0074 $\pm$ 0.0018
II	5	0.0000	0.0079	0.0060 $\pm$ 0.0022	0.0032	0.0278	0.0034	0.0090 $\pm$ 0.0026	0.0227	0.0042	0.0085 $\pm$ 0.0025
III	6	0.0000	0.0131	0.0100 $\pm$ 0.0022	0.0149	0.0045	0.0021	0.0027 $\pm$ 0.0007	0.0037	0.0040	0.0040 $\pm$ 0.0008
IV	2	0.0000	0.0000	0.0000	0.0000	0.0347	0.0021	0.0096 $\pm$ 0.0048	0.0284	0.0017	0.0079 $\pm$ 0.0039
V	2	0.0316	0.0000	0.0074 $\pm$ 0.0037	0.0000	0.0204	0.0000	0.0048 $\pm$ 0.0024	0.0224	0.0000	0.0053 $\pm$ 0.0026
Average	—	0.2480	0.0073	0.0523 $\pm$ 0.0061	0.0941	0.2226	0.0098	0.0522 $\pm$ 0.0062	0.2264	0.0093	0.0522 $\pm$ 0.0061

Values of nucleotide diversity ( $\pi$ ) were obtained according to Nei (62) with Jukes–Cantor correction for multiple substitutions; in the intergenic spacer segregating sites in indels were considered.

Overall, the deduced *kai* amino acid sequences in *N. linckia* shared 90.9% identity in the 254-residue region amplified. Walker’s motif  $A_s$  (P-loop 1, residues 91–98), reported to be important in ATP/GTP binding and circadian rhythm regulation (18, 39), is not conserved; in the copy *kaiIIId1*, occurring on the stressful SFS in EC I, polar amino acid serine was replaced by nonpolar proline. Another motif, DXXG (where X is any residue), which is essential also for circadian oscillation (39), occurs twice in the studied region (residues 113–116 and 156–159) and is not conserved in the first case (either P115S or P115A). The most variable region within residues 51–61 is likely related to desiccation stress response; substitutions there are mostly in favor of hydrophobic amino acids.

The sequence alignment of 1,268 bp showed a total of 241 segregating sites (including those in indels) or 19%, of which 119 (15.6%) were in coding regions, and 122 (24.3%) were in the intergenic spacer. A large proportion of polymorphism in the spacer referred to indels that results in variation of its length from 76 (*kaiIV*) to 499 bp (*kaiII*).

Data on nucleotide polymorphism within paralogous subfamilies of *kai* multigene family are presented in Table 2. They suggest that although the total level of within-subfamily polymorphism essentially is the same, its patterns may vary significantly between different parts of the studied region. For example, the rate of synonymous substitutions in the *kaiIII* subfamily ( $K_s = 0.0037$ ) is about six times lower as that in the *kaiII* subfamily ( $K_s = 0.0227$ ), whereas in the intergenic spacer we observe an opposing picture (0.0149 and 0.0032, respectively). Members of the *kaiIV* subfamily show no silent substitutions in the *kaiB* gene and the intergenic spacer but indicate the highest  $K_s$  level in the *kaiC* gene (0.0347). In all the subfamilies (except *kaiIII*), the level of silent substitutions in the intergenic spacer is considerably lower than in the coding regions of the *kai* cluster. This result suggests that both coding regions and the spacer are under different selective constraints. The same is referred to the different subfamilies; presumably oldest, *kaiV* has a  $K_s$  value virtually the same as the most recent *kaiII* (0.0224 and 0.0227, respectively).

A distinctive feature of polymorphism within *kai* gene subfamilies is that only singletons represent it (of course, here we do not consider the subfamilies *kaiIV* and *kaiV* made up of only two members). Moreover, there are only two mutations shared between the subfamilies (Table 3). Overall, between-subfamily variability accounts for 93% of the total variability observed in the *kai* family. These facts provide proof for the recent duplication of the genes and their primarily functional divergence.

**Codon Usage Bias and G+C Content.** The studied region in all *kai* subfamilies, as well as in the *kaiBC* region from the other *Nostocaceae*, appears to have strong codon usage bias (Table 4). In *N. linckia* the mean value of CBI was 0.473. The analysis of relative codon usage tables (data not shown) indicates that the bias is in

**Table 3. Variation within and between subfamilies of the *kai* gene family in *N. linckia***

Subfamilies	<i>kail</i>	<i>kaiII</i>	<i>kaiIII</i>	<i>kaiIV</i>	<i>kaiV</i>
<i>kail</i>	0.006 ± 0.002	130/0	137/0	131/1	147/0
<i>kaiII</i>	0.161 ± 0.014	0.007 ± 0.001	28/0	72/1	72/0
<i>kaiIII</i>	0.115 ± 0.012	0.036 ± 0.005	0.006 ± 0.001	72/0	71/0
<i>kaiIV</i>	0.094 ± 0.011	0.093 ± 0.010	0.088 ± 0.010	0.007 ± 0.003	61/0
<i>kaiV</i>	0.114 ± 0.013	0.092 ± 0.010	0.086 ± 0.010	0.078 ± 0.010	0.005 ± 0.003

Between-subfamilies diversity and its standard error are shown below the diagonal. The pairwise numbers of fixed/shared mutations are indicated above the diagonal. The diagonal shows within-subfamily diversity. The Tamura–Nei model of substitutions (63) was used for the calculations.

favor of NNT and NNA codons over NNC and NNG codons, which is expressed in the lower G+C content (0.359) in the *kaiB* gene compared with *kaiC* (0.423), especially at third codon positions (0.236 and 0.350, respectively). Estimating codon usage in complete *kaiB* and *kaiC* sequences of the closest relative *N. punctiforme* results in a CBI of 0.545 and 0.487, respectively. Based on the partial sequences of *kaiB* and *kaiC* from *N. linckia* and *N. punctiforme* and their respective values of CBI (data not shown), we could suppose rather firmly that the values of CBI for complete sequences of these genes in *N. linckia* are similar to those in *N. punctiforme*. Significantly different CBI values between *kaiB* and *kaiC* is a very interesting fact, taking into account that both these genes are regulated by a single promoter (18). In the most recent subfamilies *kaiII* and *kaiIII*, this bias is lower ( $P < 0.001$ ) than in the other evolutionarily older subfamilies but still higher than in *kaiC*.

### Discussion

The phenomenon of the *kai* multigene family is quite unique for prokaryotes. In fact, this family has resulted from duplications of a whole cluster of three genes ( $\approx 3$  kb) rather than of a single gene. We have not found data in the available literature about duplication of such a long functional region of genomic DNA in prokaryotes.

The *kai* family contains at least five distinct subfamilies. Two of them, *kaiIV* and *kaiV*, occur only in EC I and EC II, respectively. Interestingly, genes of the *kaiIV* subfamily were found only in single-copy haplotypes, whereas genes of the *kaiV* subfamily were observed in multicopy haplotypes (Table 1).

### Rapid Evolution of the *Kai* Gene Family Under the Local Environmental Stress

The *kai* multigene family is relatively large and evolutionarily recent. In *N. linckia*, we found at least 20 different functional genes and pseudogenes, and further sequencing, most likely, will reveal additional members. The absence of respective homologs in many other species of cyanobacteria, and even closely related *Nostocaceae* (e.g., *Nodularia*, *Cylindrospermum*, and *N. punctiforme*), indicates that origin and radiation of the family occurred after the speciation of *N. linckia*. A milestone question is, did the duplications occur before or after the canyons evolved? The high number of singletons, low within-subfamily and high between-subfamily diversity, the occurrence of the multiple-copy haplotypes only in the stressful SFS stations of both canyons, and low nucleotide polymorphism in the pseudogenes support the latter assumption; this gene family originated and has evolved along with the geological evolution of the microsites EC I and EC II. There are a lot of data about numerous gene families in prokaryotes (i.e., refs. 40–43) that suggest that the number of universally conserved gene families is small and thus assumes horizontal gene transfer and genome fusion as major forces in the evolution of prokaryotes (44). In the case of the *kai* gene family, we do not have evidence for the latter; instead, a stressful environment and high UV radiation probably give rise to genome instability and more replication errors, which may lead to frequent duplications.

The general areas of the canyons initially were occupied most likely by strains with a single ancestral copy of the *kai* cluster. Deepening the relief of the canyons by increasing erosion during the uplift of Mt. Carmel resulted in sharpening the interslope environ-

**Table 4. Comparative data on codon usage and G+C content in the studied region of the *kai* gene family and some other genes of *N. linckia* and other cyanobacteria**

Gene (accession number)	Promoter/ intergenic spacer*		Coding region		
	G+C	CBI	G+C2	G+C3s	G+C
<i>kai</i>					
<i>Nodularia</i> PCC 73104	0.286	0.425	0.368	0.336	0.419
<i>Cylindrospermum</i> PCC 7417	0.289	0.398	0.375	0.353	0.423
<i>N. punctiforme</i>	0.246	0.538	0.370	0.273	0.400
<i>kail</i> , <i>N. linckia</i>	0.350	0.452	0.373	0.323	0.413
<i>kaiII</i> , <i>N. linckia</i>	0.405	0.482	0.378	0.334	0.414
<i>kaiIII</i> , <i>N. linckia</i>	0.314	0.490	0.375	0.332	0.412
<i>kaiIV</i> , <i>N. linckia</i>	0.226	0.470	0.366	0.342	0.415
<i>kaiV</i> , <i>N. linckia</i>	0.321	0.458	0.374	0.320	0.411
Average in <i>kai</i> of <i>N. linckia</i>	0.368	0.473	0.374	0.330	0.413
Acetyl-CoA synthase, <i>N. linckia</i> (AY037787)	—	0.712	0.463	0.918	0.661
<i>dnaK</i> , <i>N. linckia</i> (AF388880)	0.374	0.409	0.439	0.421	0.490
<i>sodF</i> , <i>N. linckia</i> (AY055243)	0.320	0.252	0.370	0.464	0.455
Ribosomal protein L12, <i>Synechocystis</i> (X53178) <sup>†</sup>	0.368	0.693	0.398	0.408	0.484
<i>trpA</i> , <i>N. punctiforme</i> (AF288131)	0.478	0.405	0.439	0.409	0.487

G + C2 – GC content at the second codon positions; G + C3s – GC content at third (synonymous) codon positions; G + C – GC content in the genomic region.

\*In *kai* gene.

<sup>†</sup>Ref. 64.

**Table 5. Maximum likelihood estimates of selection intensity and mutation rate in the *kai* gene family of *N. linckia* based on the Poisson random field model**

Sites	No. of sites	$u$	$\mu_g$	$N_e^*$	$\gamma$	LL <sub>exp</sub>	LL <sub>obs</sub>	$P$ , one d.f.
Synonymous	178 sites	5.84 ± 1.44	$3.28 \times 10^{-2} \pm 0.81 \times 10^{-2}$	$7.81 \times 10^6 \pm 1.93 \times 10^6$	2.84 ± 3.84	-38.799	-35.342	0.0085
Nonsynonymous	254 codons	9.43 ± 3.56	n.a.	n.a.	-3.76 ± 2.46	-19.984	-13.672	0.0004

$u$ , total mutation rate (scaled to  $N_e$ );  $\mu_g$ , mutation rate per site per generation (scaled to  $N_e$ );  $\gamma$ , selection coefficient (scaled to  $N_e$ ); LL<sub>exp</sub>, expected log likelihood (provided  $\gamma = 0$ ); LL<sub>obs</sub>, observed log likelihood (for the obtained  $\gamma$ ); n.a., not applied; d.f., degree of freedom.

\*Estimated with  $\mu_g = 4.20 \times 10^{-9}$  (assuming per site per year  $\mu = 4.20 \times 10^{-8}$  and 10 generations per year on average).

mental conditions at a microscale and, consequently, in further single-copy *kai* cluster evolution. At some period, different patterns of geological development of the canyons brought about unequal rates of molecular evolution in the *kai* subfamilies and, consequently, different times for the first cluster's duplications in the canyons (Fig. 1). Despite the potential for a high migration rate between the slopes and the canyons (which is confirmed by the common occurrence of haplotypes with the invariable copy *kaila*), genetic diversity within the gene family is determined largely by the local contrasting microclimatic factors in the microsites. This is confirmed by the fact that all the haplotypes are unique to the canyons in which they occur. Furthermore, they are not observed even in the lowest stations, SFS3, of the same African slopes of the canyons. The unique nonsense mutations in all observed pseudogenes also support the hypothesis that they evolved independently. Additionally, if one supposes that the multiple-copy haplotypes were introduced into the canyons and then evolved further, one should expect to find such haplotypes on the two opposite slopes, but this was not the case.

#### Driving Forces of *Kai* Gene Family Evolution: Selection vs. Neutrality.

According to the theory elaborated thus far, there are three potential ways of the further evolution of duplicated genes: (i) one of the copies maintains the original function, whereas another may accumulate deleterious mutations and become nonfunctional (nonfunctionalization), (ii) one copy acquires a novel, advantageous function and becomes rapidly fixed in a population, and another copy still keeps the original function (neofunctionalization), and (iii) duplication may lead to the partitioning of original functions such that both new copies have reduced efficiency compared with the ancestral gene (subfunctionalization; refs. 45–49).

In the origin and evolution of the *kai* multigene family, one can clearly determine the case of nonfunctionalization (pseudogenes  $\psi$ *kaila*,  $\psi$ *kailg*, and  $\psi$ *kailIIg*). On the other hand, neofunctionalization and subfunctionalization seem not to have occurred during this evolution. The former demands sufficiently large divergence between the copies to attain a new function by one of them. For the *kailI* and *kailIII* subfamilies, it is hardly probable because of the short period elapsed since the duplication. In turn, subfunctionalization implies lower fitness of each newly arising gene, but it is not the case for the *kai* subfamilies, because almost all multiple-copy haplotypes (except haplotype D) contain virtually the same *kaila* allele, which occurs in the single-copy strains ubiquitous in all locations and, consequently, have at least the same or similar fitness. Therefore, there is no evidence for the decreased fitness of the newly duplicated genes. Instead, our data suggest that these duplications may have adaptive significance and most likely increase fitness. A similar assumption, that gene duplications (at least in enteric bacteria) are quite rapid and perhaps the first response to strong selection, was made in the early *Escherichia coli* literature (50). We propose to consider such a way of evolution of duplicated genes as *superfunctionalization* (*reinforcement*).

Indeed, evolution generally should increase fitness of organisms to ensure their survival. In the ECs, rise in the daily abundance of UV and visible radiation on the SFS, compared with the NFS, should favor temporal intracellular adaptation to the local day/night conditions to optimize programming for the

cellular photochemically dependent processes. This process led to either the adaptive evolution of a single copy of the *kai* cluster or the appearance of multiple copies. Recently Haack and Roth (51) showed that in populations of enteric bacteria, spontaneous tandem chromosomal duplications themselves are fairly common; most of them are unstable and lost, but some may confer a selective advantage under specific conditions. Our data on patterns of synonymous and nonsynonymous substitutions in the different domains of the studied region of the *kai* cluster (Table 1) provide further evidence for the predominance of selective factors operating to shape the observed polymorphism. The most recently duplicated subfamilies *kailI* and *kailIII* have only replacement substitutions fixed in the *kaiB* gene, providing clear evidence for positive selection. Nonsynonymous mutations arising *de novo* in this domain of the subfamilies after the duplication conferred new functional features to the gene apparently had adaptive significance and thus became maintained in the population. For instance, replacing polar amino acid serine by nonpolar proline in ATP/GTP binding site of *kailIII**d1* likely is adaptive to desiccation, because proline was shown to be related to drought-stress response (52–54). The substitutions in the 51–61 amino acid region probably have similar adaptive significance, because they are generally in favor of hydrophobic amino acids. The strains with the multiple gene copies seem to have a selective advantage under harsh conditions merely because nonsense mutations in one or even two copies of the cluster (as in haplotypes E and G, Fig. 1) do not result in the death of the mutant. By contrast, polymorphism of *kailI*, *kailIV*, and *kailV* subfamilies has already acquired adaptive significance to the specific local environmental conditions, such that most of the replacements become deleterious and, therefore are eliminated by purifying selection. Thus, the low within-subfamily nucleotide polymorphism is governed by purifying selection, whereas the high haplotype diversity is maintained by diversifying and stabilizing selections.

Comparison of codon usage in *kaiBC* genes to that in some highly and moderately expressed genes (Table 4) suggests that the *kai* genes are highly expressed. Of the two *kai* genes, *kaiB* has a stronger codon usage bias and probably higher expressivity than *kaiC*. Codon usage patterns usually are well conserved during evolution and confer selective constraints to the rate of synonymous substitutions (55). In the *kai* gene family, despite the high between-subfamily rate of synonymous substitutions, the within-subfamily rate is low (Table 2), and both are constrained to maintain the low GC content. The observed codon bias in the *kai* genes in favor of A/T-ended codons may be related to an overall high A+T content of the genome.

The maximum likelihood estimates using the Poisson random field model (33) give further strong support for selection as a principle driving force of *kai* gene family evolution (Table 5). They suggest that selection operates against disfavored amino acids ( $\gamma = -3.76$ ) and in favor of synonymous substitutions ( $\gamma = 2.84$ ), which confirms the general prevalence of purifying selection in the whole studied *kaiBC* region. However, in the particular *kai* regions and subfamilies, different types of selection may be predominant as is, for example, a case of positive selection described above for *kaiB* gene of *kailI* and *kailIII* subfamilies.

The scaled per site per generation mutation rate obtained by using  $\theta$  is consistent with that computed with maximum likelihood estimates ( $\mu_g$ , Table 5). Actually, given synonymous  $\theta = 0.162$  and, on average, 10 generations of *Nostoc* per year (56) yields  $\mu_g = 0.81 \times 10^{-2}$ , which is only four times less than the value obtained with maximum likelihood estimates. However, under conditions on the SFS of both canyons, the life cycle of cyanobacteria may increase considerably in length. The longer generation may take place on the one hand because of generally favorable conditions in the rainy season (October–May) when vegetative stages of *Nostoc* prevail, and on the other hand, during the dry season (May–October) because of longer intervals between the spores' production and germination (56). Therefore, *Nostoc* on the SFS of both canyons most likely has 4–5 generations per year. In such a case, values of  $\mu_g$  calculated with the two mentioned methods become virtually the same. Such consistency of the results obtained by different approaches strongly confirms our assumptions about the patterns of the *kai* gene family evolution.

Do neutral factors operate to shape polymorphism in the *kai* genes? According to the neutral theory, neutral genetic diversity ( $\theta$ ) in a haploid population is governed by the effective population size ( $N_e$ ) and mutation rate ( $\mu$ ),  $\theta = 2N_e\mu$  (17). Thus, large population size and high mutation rate should increase the level of polymorphism. As stated above, the neutral mutation rate in the *kai* genes from *N. linckia* in the stressful SFS stations of the ECs was determined to be  $4.20 \times 10^{-8}$  substitutions per site per year, which is almost 200 times higher than in the *kai* genes of other cyanobacteria. It also is almost 10–20 times higher than the rate of neutral substitutions in endosymbiotic *E. coli* and *Buchnera* (57). The effective population size of *N. linckia* from those slopes is estimated to be  $7.81 \times 10^6 \pm 1.93 \times 10^6$  (Table 5). It is almost 2 orders of magnitude lower than the estimated  $N_e$  for *E. coli* (33) and, most likely, significantly lower than the  $N_e$  for *N. linckia* from the NFS of the canyons. The smaller effective population size thus may promote faster fixation of neutral and slightly advantageous sub-

stitutions. But this factor apparently plays a minor role in the evolution of the *kai* gene family.

An intriguing question concerns the level of polymorphism in the *kaiBC* intergenic spacer. In all the subfamilies except *kaiIII*, it indicates significantly lower nucleotide diversity than in the coding regions (Table 2), although it was expected to be at least as high as the level of synonymous substitutions. There are numerous data that suggest that intergenic spacers and 5'-flanking sequences may play an important role in the regulation of transcription (58–60). Therefore, although the intergenic spacer was reported to have no functional significance in the *kai* cluster (18), our data on its nucleotide polymorphism may suggest differently, at least in some *kai* subfamilies.

The long-term study of the evolutionary patterns under an acute microclimatic stress in EC I revealed that, across phylogeny, permanent ecological stress usually results in a higher mutation rate and an increase in genetic polymorphism (2, 3, 61). This higher polymorphism in the cyanobacteria may arise, in addition to other factors, from the multiple gene duplication events, which probably are one of the ways that cyanobacteria adapt to the extreme and fluctuating environment. Under these conditions, evolution of functionally important genes such as those regulating circadian rhythms is governed primarily by various types of selection rather than by purely neutral factors. The described example of extremely rapid evolution of the clock gene family controlling circadian rhythmicity, one of the fundamental features in most organisms, helps to better understand the processes underlying adaptation by natural selection to environmental stress.

We thank Mrs. T. Krugman and Dr. N. Satish for help in collecting cyanobacterial strains and technical assistance in DNA isolation. We also thank Prof. O. Savolainen (University of Oulu), Dr. C. Vogl (University of Munich), and Prof. D. Hartl (Harvard University) for valuable comments on the manuscript. This work was supported by the Israeli National Science Foundation Grant 98-04-4963, the Israeli Discount Bank Chair of Evolutionary Biology, and the Ancell-Teicher Research Foundation for Genetics and Molecular Evolution.

- Nevo, E. (1995) *Proc. R. Soc. London B* **262**, 149–155.
- Nevo, E. (1997) *Theor. Popul. Biol.* **52**, 231–243.
- Nevo, E. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 6233–6240.
- Whitton, B. A. (1987) in *Survival and Dormancy of Microorganisms*, ed. Hennis, Y. (Wiley, New York), pp. 109–167.
- Clarke, A. K., Soitamo, A., Gustafsson, P. & Oquist, G. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 9973–9977.
- Kulkarni, R. D. & Golden, S. S. (1994) *J. Bacteriol.* **176**, 959–965.
- Erdmann, N., Effmert, U., Fulda, S. & Oheim, S. (1997) *Curr. Microbiol.* **35**, 348–355.
- Tanaka, N. & Nakamoto, H. (1999) *FEBS Lett.* **458**, 117–123.
- Vass, I., Kirilovsky, D., Perewoska, I., Mate, Z., Nagy, F. & Etienne, A. L. (2000) *Eur. J. Biochem.* **267**, 2640–2648.
- Van Valen, L. (1965) *Am. Nat.* **99**, 377–390.
- Soulé, M. & Stewart, B. R. (1970) *Am. Nat.* **104**, 85–97.
- Nevo, E. (1988) *Evol. Biol.* **23**, 217–246.
- Nevo, E. (1998) *J. Exp. Zool.* **282**, 95–119.
- Vinogradova, O. N., Kovalenko, O. V., Wasser, S. P., Nevo, E., Kislova, O. A. & Belikova, O. A. (2000) *Algologia* **5**, 46–55.
- Satish, N., Krugman, T., Vinogradova, O. N., Nevo, E. & Kashi, Y. (2001) *Microb. Ecol.* **42**, 306–316.
- Krugman, T., Satish, N., Vinogradova, O. N., Beharav, A., Kashi, Y. & Nevo, E. (2001) *Evol. Ecol. Res.* **3**, 899–915.
- Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.).
- Ishiura, M., Kutsuna, S., Aoki, S., Iwasaki, H., Andersson, C. R., Tanabe, A., Golden, S. S., Johnson, C. H. & Kondo, T. (1998) *Science* **281**, 1519–1523.
- Lorne, J., Scheffer, J., Lee, A., Painter, M. & Miao, V. P. (2000) *FEMS Microbiol. Lett.* **189**, 129–133.
- Johnson, C. H., Golden, S. S. & Kondo, T. (1998) *Trends Microbiol.* **6**, 407–410.
- Taniguchi, Y., Yamaguchi, A., Hijikata, A., Iwasaki, H., Kamagata, K., Ishiura, M., Go, M. & Kondo, T. (2001) *FEBS Lett.* **496**, 86–90.
- Xu, Y., Mori, T. & Johnson, C. H. (2000) *EMBO J.* **19**, 3349–3357.
- Chen, Y. B., Dominic, B., Mellon, M. T. & Zehr, J. P. (1998) *J. Bacteriol.* **180**, 3598–3605.
- Mori, T., Binder, B. & Johnson, C. H. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 10183–10188.
- Huang, T.-C., Chen, H.-M., Pen, S.-Y. & Chen, T. H. (1994) *Planta* **193**, 131–136.
- Johnson, C. H. & Golden, S. S. (1999) *Annu. Rev. Microbiol.* **53**, 389–409.
- Ouyang, Y., Andersson, C. R., Kondo, T., Golden, S. S. & Johnson, C. H. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 8660–8664.
- Golden, S. S., Brusslan, J. & Haselkorn, R. (1987) *Methods Enzymol.* **153**, 215–231.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
- Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3**, 418–426.
- Morton, B. R. (1993) *J. Mol. Evol.* **37**, 273–280.
- Rozas, J. & Rozas, R. (1999) *Bioinformatics* **15**, 174–175.
- Hartl, D. L., Moriyama, E. N. & Sawyer, S. A. (1994) *Genetics* **138**, 227–234.
- Felsenstein, J. (1981) *J. Mol. Evol.* **17**, 368–376.
- Xia, X. (2000) *Data Analysis in Molecular Biology and Evolution* (Kluwer, Boston).
- Tyler, S. A. & Barghoorn, E. S. (1954) *Science* **119**, 606–608.
- Cloud, P. E. (1965) *Science* **148**, 27–35.
- Karcz, Y. (1959) *Bull. Res. Council. Isr.* **8**, 119–130.
- Nishiwaki, T., Iwasaki, H., Ishiura, M. & Kondo, T. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 495–499.
- Hill, K. E., Marchesi, J. R. & Weightman, A. J. (1999) *J. Bacteriol.* **181**, 2535–2547.
- Makarova, K. S., Aravind, L., Galperin, M. Y., Grishin, N. V., Tatusov, R. L., Wolf, Y. I. & Koonin, E. V. (1999) *Genome Res.* **9**, 608–628.
- Yanai, I., Camacho, C. J. & DeLisi, C. (2000) *Phys. Rev. Lett.* **85**, 2641–2644.
- Carlyon, J. F., Roberts, D. M. & Marconi, R. T. (2000) *Microb. Pathog.* **28**, 89–105.
- Koonin, E. V. & Galperin, M. Y. (1997) *Curr. Opin. Genet. Dev.* **7**, 757–763.
- Ohno, S. (1970) *Evolution by Gene Duplication* (Springer, Berlin).
- Walsh, J. B. (1985) *Genetics* **110**, 345–364.
- Clark, A. G. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 2950–2954.
- Sidow, A. (1996) *Curr. Opin. Genet. Dev.* **6**, 715–722.
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. & Postlethwait, J. (1999) *Genetics* **151**, 1531–1545.
- Langridge, J. (1969) *Mol. Gen. Genet.* **105**, 74–83.
- Haack, K. R. & Roth, J. R. (1995) *Genetics* **141**, 1245–1252.
- Yoshida, Y., Kiyosue, T., Nakashima, K., Yamaguchi-Shinozaki, K. & Shinozaki, K. (1997) *Plant Cell Physiol.* **38**, 1095–1102.
- Harrak, H., Chamberland, H., Plante, M., Bellemare, G., Lafontaine, J. G. & Tabaeizadeh, Z. (1999) *Plant Physiol.* **121**, 557–564.
- Menke, U., Renault, N. & Mueller-Roeber, B. (2000) *Plant Physiol.* **122**, 677–686.
- Ikemura, T. (1985) *Mol. Biol. Evol.* **2**, 13–34.
- Kondratieva, N. V., Posudin, Yu. I. & Belikova, O. A. (1987) *Ukr. Bot. J.* **44**, 21–25.
- Ochman, H., Elwyn, S. & Moran, N. A. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 12638–12643.
- Busby, S. J. & Reeder, R. H. (1983) *Cell* **34**, 989–996.
- Flavell, R. B., O'Dell, M., Sharp, P., Nevo, E. & Beiles, A. (1986) *Mol. Biol. Evol.* **3**, 547–558.
- Lang, W. H., Morrow, B. E., Ju, Q., Warner, J. R. & Reeder, R. H. (1994) *Cell* **79**, 527–534.
- Lamb, B. C., Saleem, M., Scott, W., Thapa, N. & Nevo, E. (1998) *Genetics* **149**, 87–99.
- Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).
- Tamura, K. & Nei, M. (1993) *Mol. Biol. Evol.* **10**, 512–526.
- Sibold, C. & Subramanian, A. R. (1990) *Biochim. Biophys. Acta* **1050**, 61–68.