Homing endonucleases residing within inteins: evolutionary puzzles awaiting genetic solutions

Adi Barzel*, Adit Naor*, Eyal Privman†, Martin Kupiec* and Uri Gophna*[1]

*Department of Molecular Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel, 69978, and
†Department of Cell Biology and Immunology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel, 69978

Abstract

Inteins are selfish genetic elements that disrupt the sequence of protein-coding genes and are excised post-translationally. Most inteins also contain a HEN (homing endonuclease) domain, which is important for their horizontal transmission. The present review focuses on the evolution of inteins and their nested HENs, and highlights several unsolved questions that could benefit from molecular genetic approaches. Such approaches can be well carried out in halophilic archaea, which are naturally intein-rich and have highly developed genetic tools for their study. In particular, the fitness effects of habouring an intein/HEN can be tested in direct competition assays, providing additional insights that will improve current evolutionary models.

HENs (homing endonucleases)

HENs are a large and diverse class of selfish elements found in archaea, bacteria and lower eukaryotes, and their respective viruses. HENs recognize and cleave specifically long target sequences (12–40 bp) that typically occur only once in a given genome [1,2]. Remarkably, HENs are found almost exclusively within self-splicing selfish elements, namely introns, which excise themselves at the mRNA level, and inteins, which splice out of the protein product [3] (Figure 1a). HENs contribute to the horizontal propagation of these selfish elements into intron-less or intein-less alleles, by cleaving the vacant allele to induce homologous recombination or reverse transcription (Figure 1b). Since HENs within group I introns were reviewed recently [4], we focus on HENs occupying inteins, examine the interesting evolutionary questions they bring up, and propose ways in which molecular biology experiments can help resolve undecided evolutionary conundrums.

Mutualism between selfish elements

Although it is obvious that inteins can profit from horizontal invasion, how and why the HEN benefits from its association with the intein is less obvious. Since HENs typically confer no selective advantage upon their host organism, they can only persist as long as the rate of their dissemination surpasses the rate of their degeneration via genetic drift or counterselection. The rate of dissemination is in turn dependent on the availability of homing targets, which are also subjects of mutation and selection. If cleavage by the HEN is by any degree toxic to the host, mutations in the target site may be selected that prevent it from being recognized or cleaved.

Even if the HEN is neutral with respect to the survival of the host, the sequence of the target site may drift faster than can be accommodated by the evolution of the HEN. The long-term survival of a HEN is therefore dependent on the conservation of its target site. This target site is also the eventual insertion site of the HEN, and an insertion in a conserved locus will often dramatically reduce host fitness. Conversely, a HEN will cause no harm in terms of target disruption if it inhabits an intein, since it will be spliced out, preserving the function of their hosting gene. Moreover, inteins, whether with or without a HEN, tend to occupy conserved regions of essential genes [5,6]. This is the intein's intrinsic way of avoiding degeneration, since any disabling mutation in its splicing domain will create a permanent insertion in the essential gene and will therefore be eradicated. Thus, in this association, we observe mutualism between selfish elements: the HEN benefits from being nested within a spliceable ORF (open reading frame) integrated within a conserved site (the intein), while at the same time providing that ORF with transmissibility.

How did the mutualism between selfish elements evolve?

Free-standing HENs are known to exist in bacteriophages, where they can cleave the DNA of a competing bacteriophage, but not the one carrying them, thus providing a fitness advantage when two or more phages infect a host cell [7]. How HENs became associated with inteins remains unclear. However, two recent studies in bacteriophages may have revealed the process underlying the mutualistic association of introns and HENs, which is highly likely to be relevant also to inteins. In the cyanophage S-PM2, a free-standing HEN, similar in sequence to DNA resolvases, was observed close to the *psbA* gene, which contains a group I intron lacking a nested HEN [8]. This HEN, F-CphI, was
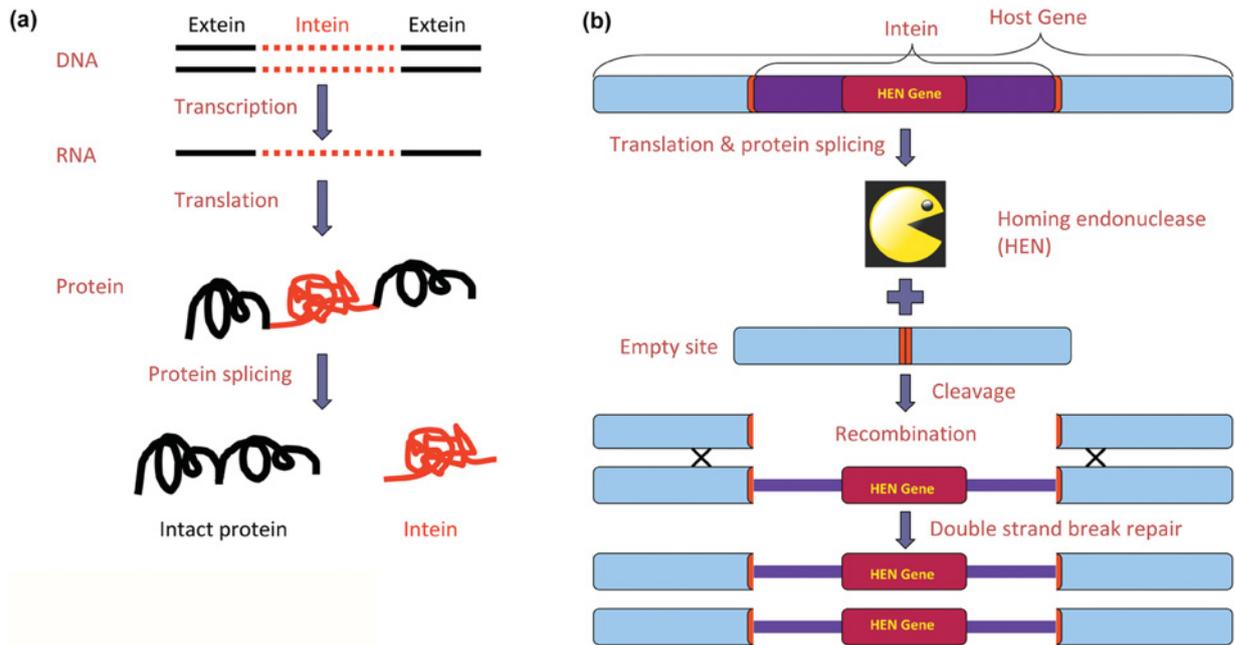
Figure 1 | Protein splicing and homing

(**a**) Inteins and protein splicing. (**b**) Target invasion by HENs.

shown to be unable to cleave the intron-containing *psbA* gene, but does cleave efficiently empty (intron-free) *psbA* alleles just a few nucleotides from the intron insertion site [9]. Furthermore, Bonocora and Shub [10] showed that, in a gene encoding DNA polymerase in T3 and T7 bacteriophages, there exists a site adjacent to an intron insertion point, which can be targeted by either an intron-nested HEN (T7) or a free-standing HEN (T3), again implying that HENs that target intron insertion sites can gradually become associated with introns. It is expected that the mutualistic relationship between the two elements was stabilized once they became physically associated.

## How do HENs avoid degeneration?

The potential mutational drift of the target site mentioned above is not the only threat for the survival of HENs. Somewhat paradoxically, the very proliferation of a HEN may sow the seeds of its own destruction: in order to prevail on an evolutionary timescale, a HEN must also overcome the danger of target site depletion as a result of its own fixation in the population. Extensive horizontal propagation of the HEN could lead to fixation, with the HEN occupying every allele of the hosting gene in the population. Once fixation is achieved, no vacancy is left for further homing. Subsequently, there will be no selective pressure favouring the conservation of the HEN, making degeneration via accumulation of mutations inevitable (Figure 2). Indeed, inteins in various species were found to encode degenerated HENs, characterized by an accumulation of asynonymous mutations [11,12] and a lack of activity in *in vitro* cleavage

assays [13,14]. Intriguingly, functionally degenerate HENs within inteins, unlike those that reside in introns, still show signs of purifying selection and maintain several highly conserved positions, perhaps suggesting that some positions within the HEN domain are also important for the protein-splicing reaction of the intein or for its overall structure [12].

One possible getaway from the fixation-degeneration deadlock is to invade new niches, in other species, via horizontal transfer (Figure 2). However, quite often, HEN phylogenies agree with those of the host organism, implying that many HENs do nevertheless persist in a specific locus through evolutionary timescales. How then do they escape the fixation-degeneration catch? The canonical model of HEN evolution proposed by Burt and Koufopanou [1], called 'the HEN cycle' or the 'homing cycle' [2], addresses this enigma by invoking the recycling of homing targets. It proclaims that, while HEN propagation acts to exhaust all available targets, there is also an opposing process, namely precise intron/intein loss, which provides the supply of fresh targets (Figure 2). Indeed, phylogenetic analyses support the idea of frequent intron/intein loss, as very often several species in a genus harbour an intron that is not found in other species of that genus. Analogously, at the subspecies level, the HEN may persist while avoiding fixation if sporadic populations exist with and without the intron, and these populations exhibit low gene flow among them.

Although one can imagine a precise deletion of an intron by homologous recombination with its own cDNA, explaining how an intein will be precisely deleted is not easy. Homologous recombination between an intein-containing

Figure 2 | The homing cycle
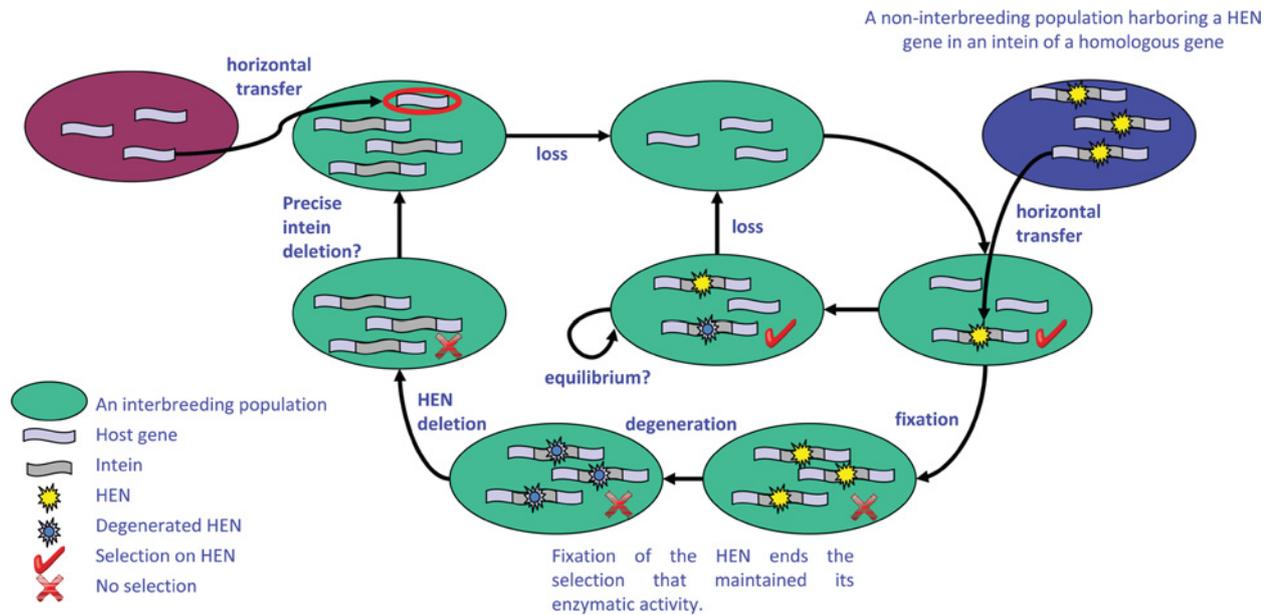
and a vacant allele will tend to result in gene conversion events in which the former is duplicated, rather than the latter. This is due to the fact that HEN activity (the creation of a double-strand break at the vacant allele) promotes gene conversion. Thus one has to posit that HEN degeneration or inactivation is a necessary step that always precedes intein deletion. However, even if such degeneration has already taken place, coming into contact with such a vacant allele should be a relatively rare event, requiring horizontal gene transfer from another population or species, and, as long as some parts of the population still have an active HEN, these events are unlikely to efficiently compete with homing. Moreover, recombination that is not triggered by HEN cleavage has a similar propensity to result in intein gain as it does to result in intein loss. Recently, both experimental data from euascomycetes species [15] and mathematical modelling [16] have shown that HENs can be maintained for very long evolutionary periods even in the absence of horizontal gene transfer. In fact, all three types of homing site can coexist in a well-mixed population assuming that the reduction in fitness resulting from carrying a dysfunctional HEN is higher than the reduction in fitness due to its functional counterpart [16].

However, currently, there is no specific evidence in support of that assumption. When the cost to host fitness of the functional HEN was actually higher than the cost of its dysfunctional counterpart, only periodic solutions were obtained [16], with typical allele frequencies that are lower than commonly observed in the natural populations surveyed to date [15]. Several alternative routes for HEN persistence have been suggested [2]. One such scenario that is in agreement with selfish element mutualism, states that inteins may be lost along with their resident HENs if they (and not the HENs as above) are detrimental to the

fitness of their hosts [15]. This toxicity cannot be remedied by mutations that interfere with efficient splicing, because the intein is usually inserted in a conserved site of a vital protein. However, if individuals that possess the vacant allele are still present in the population, the fitness cost of the intein is likely to drive the vacant allele to fixation, resulting in eventual intein loss. Nonetheless, the intein can be actively propagated by its nesting HEN, and if the adverse effect is relatively minor, a dynamic equilibrium can result, since, on one hand, intein-less hosts are more fit, but, on the other, homing allows the invasion of the intein into vacant alleles. This balance will ensure a constant supply of vacant targets, and therefore a continuous selective pressure acting to conserve the homing ability of the HEN. Thus, paradoxically, by residing within inteins that cause a slight reduction in fitness, HENs can be protected from the fixation-degeneration mechanism. Gogarten and Hilario [2] pointed out that, although this dynamic equilibrium may be too precise and fragile to be maintained within a population, a balance between decreased host fitness and increased spreading of the intein, combined with a complex population structure (i.e. several subpopulations in which the fitness disadvantage and the homing efficiency could vary) can generate long persistence without requiring a homing cycle. Furthermore, even if the homing cycle does exist, it could be operating only within specific subpopulations, whereas in the meta-population, the HEN-mobilized intein never reaches fixation [2].

Another intriguing alternative to the homing cycle is that the HEN itself may confer some selective advantage on its host. Although the ability to cut the DNA of competing viruses can surely be an asset to HEN-containing bacteriophages [17], other benefits, such as a beneficial role

in chromosomal DNA rearrangement, as described in the thermophilic archaeon *Pyrococcus kodakaraensis* [18], may be less obvious and more difficult to demonstrate. Another case that illustrates such possible mechanisms is that of the 'HO' HEN of *Saccharomyces cerevisiae*, derived from the intein VDE [VMA1 (vacuolar membrane ATPase 1)-derived endonuclease], which no longer propagates an intein but has been 'domesticated' and now mediates a new role, that of a mating-type switch, in that yeast [12].

Halophilic archaea as a model genetic system to study inteins and their HENs

Clearly, it is often difficult to decide between the different evolutionary scenarios presented above on the basis of mathematical modelling and phylogenetic reconstruction alone, no matter how well executed. Furthermore, incorporating experimental observations can help to define better the assumptions and parameters of the models. Ideally, studies of HEN evolution should incorporate population genetics studies, involving the comparison of inteins from same-species isolates [19,20] and molecular genetics studies, measuring the influence on fitness of HEN and/or intein presence *in vivo* and quantifying homing efficiency in a natural-like laboratory experiment that could involve both intra- and inter-species homing. Currently, the best example of such a genetic approach is the landmark intron study by Aagaard et al. [21], which showed that when the 23S rRNA HEN-containing intron of the archaeal hyperthermophile *Desulfurococcus mobilis* was artificially introduced into another hyperthermophilic archaeon, *Sulfolobus acidocaldarius*, it quickly spread through the *S. acidocaldarius* population, presumably by archaeal mating [22–24] or conjugation [25] and intron-containing cells obtained a fitness advantage over their intron-free counterparts. Such an approach could easily be extended in intein research into a more natural-like experimental setting, where HEN genes are studied in their organisms of origin.

We propose that halophilic archaea (haloarchaea), which are easier to grow and manipulate than thermophiles, represent an ideal prokaryotic model system for the study of inteins/HENs due to several unique properties as follows.

(i) They harbour a multitude of inteins. Archaea in general and haloarchaea in particular tend to contain many inteins. For example, the square archaeon *Haloquadratum walsbyi* contains 14 putative inteins according to the intein database InBase [26]. Haloarchaea often harbour different inteins in the same gene. The gene encoding DNA polymerase B in haloarchaea (*polB*), for example, can contain at least four different inteins, and at least one species (*H. walsbyi*) has inteins occupying three possible locations [27]. An intein is occasionally conserved both in terms of the position of insertion and in terms of sequence similarity, across relatively remote haloarchaeal species. For example, the *polB* intein of *Haloferax volcanii* is homologous with one of the *polB* inteins in *H. walsbyi*.

(ii) Haloarchaea undergo horizontal gene transfer by the well-documented and highly efficient cell mating process [22–24], which can allow DNA exchange within and even between species and therefore can facilitate the spread of a HEN to empty alleles within a population. Thus one can potentially monitor the dynamic of homing in real time.

(iii) Haloarchaea, in particular *H. volcanii* and *Halobacterium salinarum*, have become the most extensively used model organisms for molecular genetic studies of archaea because of the wide range of available genetic tools [28], such as shuttle vectors [29,30], the ability to manipulate their DNA by transformation [31,32], the existence of a strongly inducible promoter [33], as well as a targeted allele-exchange and gene-knockout system [34]. Finally, haloarchaea can also be harnessed to identify additional functions of HENs, which can be studied relatively simply, because host proteins can be engineered to be affinity-tagged, facilitating co-immunopurification experiments.

One can envisage several ways to use molecular genetics for the study of the evolutionary questions outlined above, by performing fairly straightforward experiments. (i) An intein-containing gene can be cured of its intein (or just the HEN domain) and compared with its wild-type counterpart, both in terms of individual growth/fitness and in direct competition experiments. (ii) A new intein can be introduced into a currently vacant site and the effects similarly examined. (iii) The invasion of a HEN-containing intein in a formerly intein-less population, can be monitored, by introducing a small number of intein/HEN-positive cells into a large intein-free population, then taking samples periodically and quantifying the fraction of intein-containing cells in the population (e.g. by quantitative real-time PCR). We have recently performed the first experiment (i), and showed that, under laboratory conditions, curing of an intein/HEN did not result in any change in fitness [35]

To date, intein/HEN research has already supplied many surprises, some of which have found industrial applications such as split inteins [36] or engineered HENs for gene therapy [37]. The future of intein/HEN research lies in combining multiple approaches including mathematical modelling, genomics, population genetics and molecular genetics. One can easily envision quantitative parameters derived from large population genetics surveys as well as molecular experiments being fed back into the existing mathematical models, increasing their predictive power. Conversely, new models and genomic observations will yield new hypotheses that need to be tested. Only by integrating multiple disciplines in the study of inteins/HENs could we hope to unravel the complex evolutionary puzzles that they are.

Funding

References

1 Burt, A. and Koufopanou, V. (2004) Homing endonuclease genes: the rise and fall and rise again of a selfish element. Curr. Opin. Genet. Dev. **14**, 609–615

2 Gogarten, J.P. and Hilario, E. (2006) Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. BMC Evol. Biol. **6**, 94

3 Pietrokovski, S. (2001) Intein spread and extinction in evolution. Trends Genet. **17**, 465–472

4 Nielsen, H. and Johansen, S.D. (2009) Group I introns: moving in new directions. RNA Biol. **6**, 375–383

5 Edgell, D.R., Stanger, M.J. and Belfort, M. (2004) Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. J. Mol. Biol. **343**, 1231–1241

6 Swithers, K.S., Senejani, A.G., Fournier, G.P. and Gogarten, J.P. (2009) Conservation of intron and intein insertion sites: implications for life histories of parasitic genetic elements. BMC Evol. Biol. **9**, 303

7 Edgell, D.R., Fast, N.M. and Doolittle, W.F. (1996) Selfish DNA: the best defense is a good offense. Curr. Biol. **6**, 385–388

8 Mann, N.H., Clokie, M.R., Millard, A., Cook, A., Wilson, W.H., Wheatley, P.J., Letarov, A. and Krisch, H.M. (2005) The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. J. Bacteriol. **187**, 3188–3200

9 Zeng, Q., Bonocora, R.P. and Shub, D.A. (2009) A free-standing homing endonuclease targets an intron insertion site in the *psbA* gene of cyanophages. Curr. Biol. **19**, 218–222

10 Bonocora, R.P. and Shub, D.A. (2009) A likely pathway for formation of mobile group I introns. Curr. Biol. **19**, 223–228

11 Okuda, Y., Sasaki, D., Nogami, S., Kaneko, Y., Ohya, Y. and Anraku, Y. (2003) Occurrence, horizontal transfer and degeneration of VDE intein family in Saccharomycete yeasts. Yeast **20**, 563–573

12 Koufopanou, V. and Burt, A. (2005) Degeneration and domestication of a selfish gene in yeast: molecular evolution versus site-directed mutagenesis. Mol. Biol. Evol. **22**, 1535–1538

13 Saves, I., Morlot, C., Thion, L., Rolland, J.L., Dietrich, J. and Masson, J.M. (2002) Investigating the endonuclease activity of four *Pyrococcus abyssi* inteins. Nucleic Acids Res. **30**, 4158–4165

14 Posey, K.L., Koufopanou, V., Burt, A. and Gimble, F.S. (2004) Evolution of divergent DNA recognition specificities in VDE homing endonucleases from two yeast species. Nucleic Acids Res. **32**, 3947–3956

15 Butler, M.I., Gray, J., Goodwin, T.J. and Poulter, R.T. (2006) The distribution and evolutionary history of the PRP8 intein. BMC Evol. Biol. **6**, 42

16 Yahara, K., Fukuyo, M., Sasaki, A. and Kobayashi, I. (2009) Evolutionary maintenance of selfish homing endonuclease genes in the absence of horizontal transfer. Proc. Natl. Acad. Sci. U.S.A. **106**, 18861–18866

17 Goodrich-Blair, H. and Shub, D.A. (1996) Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. Cell **84**, 211–221

18 Nishioka, M., Fujiwara, S., Takagi, M. and Imanaka, T. (1998) Characterization of two intein homing endonucleases encoded in the DNA polymerase gene of *Pyrococcus kodakaraensis* strain KOD1. Nucleic Acids Res. **26**, 4409–4412

19 Wikmark, O.G., Einvik, C., De Jonckheere, J.F. and Johansen, S.D. (2006) Short-term sequence evolution and vertical inheritance of the *Naegleria* twin-ribozyme group I intron. BMC Evol. Biol. **6**, 39

20 Nomura, N., Morinaga, Y., Kogishi, T., Kim, E.J., Sako, Y. and Uchida, A. (2002) Heterogeneous yet similar introns reside in identical positions of the rRNA genes in natural isolates of the archaeon *Aeropyrum pernix*. Gene **295**, 43–50

21 Aagaard, C., Dalgaard, J.Z. and Garrett, R.A. (1995) Intercellular mobility and homing of an archaeal rDNA intron confers a selective advantage over intron− cells of *Sulfolobus acidocaldarius*. Proc. Natl. Acad. Sci. U.S.A. **92**, 12285–12289

22 Mevarech, M. and Werczberger, R. (1985) Genetic transfer in *Halobacterium volcanii*. J. Bacteriol. **162**, 461–462

23 Rosenshine, I. and Mevarech, M. (1991) The kinetic of the genetic exchange process in *Halobacterium volcanii* mating. In General and Applied Aspects of Halophilic Microorganisms (Rodriguez-Valera, F., ed.), pp. 265–270, Plenum Press, New York

24 Ortenberg, R., Tchelet, R. and Mevarech, M. (1999) A model for the genetic exchange system of the extremely halophilic archaeon *Haloferax volcanii*. Microbiology and Biogeochemistry of Hypersaline Environments (Oren, A., ed.), pp. 331–338, CRC Press, Boca Raton

25 Stedman, K.M., She, Q., Phan, H., Holz, I., Singh, H., Prangishvili, D., Garrett, R. and Zillig, W. (2000) pING family of conjugative plasmids from the extremely thermophilic archaeon *Sulfolobus islandicus*: insights into recombination and conjugation in Crenarchaeota. J. Bacteriol. **182**, 7014–7020

26 Perler, F.B. (2002) InBase: the Intein Database. Nucleic Acids Res. **30**, 383–384

27 MacNeill, S.A. (2009) The haloarchaeal chromosome replication machinery. Biochem. Soc. Trans. **37**, 108–113

28 Allers, T. and Mevarech, M. (2005) Archaeal genetics: the third way. Nat. Rev. Genet. **6**, 58–73

29 Holmes, M., Pfeifer, F. and Dyall-Smith, M. (1994) Improved shuttle vectors for *Haloferax volcanii* including a dual-resistance plasmid. Gene **146**, 117–121

30 Holmes, M.L., Nuttall, S.D. and Dyall-Smith, M.L. (1991) Construction and use of halobacterial shuttle vectors and further studies on *Haloferax* DNA gyrase. J. Bacteriol. **173**, 3807–3813

31 Cline, S.W., Lam, W.L., Charlebois, R.L., Schalkwyk, L.C. and Doolittle, W.F. (1989) Transformation methods for halophilic archaebacteria. Can. J. Microbiol. **35**, 148–152

32 Cohen, A., Lam, W.L., Charlebois, R.L., Doolittle, W.F. and Schalkwyk, L.C. (1992) Localizing genes on the map of the genome of *Haloferax volcanii*, one of the Archaea. Proc. Natl. Acad. Sci. U.S.A. **89**, 1602–1606

33 Large, A., Stamme, C., Lange, C., Duan, Z., Allers, T., Soppa, J. and Lund, P.A. (2007) Characterization of a tightly controlled promoter of the halophilic archaeon *Haloferax volcanii* and its use in the analysis of the essential *cct1* gene. Mol. Microbiol. **66**, 1092–1106

34 Bitan-Banin, G., Ortenberg, R. and Mevarech, M. (2003) Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene. J. Bacteriol. **185**, 772–778

35 Naor, A., Lazary, R., Barzel, A., Papke, R.T. and Gophna, U. (2011) *In vivo* characterization of the homing endonuclease within the *polB* gene in the halophilic archaeon *Haloferax volcanii*. PLoS ONE, in the press

36 Sun, W., Yang, J. and Liu, X.Q. (2004) Synthetic two-piece and three-piece split inteins for protein *trans*-splicing. J. Biol. Chem. **279**, 35281–35286

37 Paques, F. and Duchateau, P. (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. Curr. Gene Ther. **7**, 49–66