Gene 877 (2023) 147533

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Mimicking the last step of gene elongation: hints from the bacterial *hisF* gene

Sara Del Duca [a,b], Alberto Vassallo [c], Giulia Semenzato [a], Renato Fani [a,*]

[a] Department of Biology, University of Florence, Via Madonna del Piano 6, 50019 Sesto Fiorentino (FI), Italy
[b] Research Centre for Agriculture and Environment, Council for Agricultural Research and Economics (CREA-AA), Via di Lanciola 12/A, 50125 Cascine del Riccio (FI), Italy
[c] School of Biosciences and Veterinary Medicine, University of Camerino, Via Gentile III Da Varano 1, 62032 Camerino (MC) Italy

ARTICLE INFO

Edited by: Justin Blumenstiel

Keywords:
Evolutionary genetics
Molecular evolution
Histidine biosynthesis
Escherichia coli
Selective pressure

ABSTRACT

Gene elongation consists in an in-tandem duplication of a gene and divergence and fusion of the two copies, resulting in a gene constituted by two divergent paralogous modules. Many present-day proteins show internal repeats of amino acid sequences, generated by gene elongation events; however, gene elongation is still a poorly studied evolutionary molecular mechanism. The most documented case is that of the histidine biosynthetic genes *hisA* and *hisF*, which derive from the gene elongation of an ancestral gene half the size of the extant ones.

The aim of this work was to experimentally simulate the possible last step of the gene elongation event occurred during *hisF* gene evolution under selective pressure conditions.

Azospirillum brasilense hisF gene, carrying a single nucleotide mutation that generates a stop codon between the two halves of the gene, was used to transform the histidine-auxotrophic *Escherichia coli* strain FB182 (*hisF892*). The transformed strain was subjected to selective pressure (i.e., low concentration/absence of histidine in the growth medium) and the obtained mutants were characterized.

The restoration of prototrophy was strongly dependent on the time of incubation and on the strength of the selective pressure. The mutations involved the introduced stop codon with a single base substitution and none of the mutants restored the wild-type codon. Possible correlations between the different mutations and i) *E. coli* codon usage, ii) three-dimensional structures of the mutated HisF proteins, and iii) growth ability of the mutants were investigated. On the contrary, when the experiment was repeated by mutating a more conserved codon, only a synonymous substitution was obtained.

Thus, experiments performed in this study allowed to mimic a possible gene elongation event occurred during the evolution of *hisF* gene, evidencing the ability of bacterial cells to modify their genome in short times under selective conditions.

1. Introduction

Living organisms are highly ordered and complex systems, constituted by a multitude of biomolecular components that interact with each other in a concerted manner. This complex network ensures the functioning, survival, and reproduction of cells (Caetano-Anollés et al., 2009). Metabolic pathways are a characteristic feature of all known cells and represent the most primitive kind of organization, in which several reactions operate in sequence (Waley, 1969).

Different molecular mechanisms may have been responsible for the expansion and shaping of early genomes and metabolic abilities during evolution; these include i) internal mechanisms such as substitutions, insertions, deletions, gene elongation, duplication and/or fusion, modular assembly of new proteins, and ii) external mechanisms like cell fusion (synology), horizontal gene transfer (xenology) (Fondi et al., 2009; Levasseur and Pontarotti, 2011), and vesiculation (Faddetta et al., 2022). The spread of metabolic routes among microorganisms might have occurred through xenology and/or synology events, which could

Abbreviations: A, adenine; C, cytosine; G, guanine; Gln, glutamine; Glu, glutamate; His, histidine; IPTG, isopropyl β-D-thiogalactoside; LB, lysogeny broth medium; Leu, leucine; LUCA, Last Universal Common Ancestor; Lys, lysine; MMD, minimal medium Davis; OD_{600}, optical density 600 nm; Ser, serine; T, thymine; tRNA, transfer RNA; Tyr, tyrosine; wt, wild-type.

* Corresponding author.
E-mail address: renato.fani@unifi.it (R. Fani).

https://doi.org/10.1016/j.gene.2023.147533
Received 23 February 2023; Received in revised form 5 May 2023; Accepted 31 May 2023
Available online 4 June 2023

have been facilitated by the absence of a cell wall in primordial cells (Fani and Fondi, 2009). The acquisition of new genetic material could have been achieved also by the internal duplication of DNA regions, whilst genes and genomes shaping leading to new functions (generally gained through evolutionary divergence) could have been determined by the other cited internal molecular mechanisms (Fani and Fondi, 2009).

Concerning the internal evolutionary mechanisms, gene duplication is among the oldest and perhaps one of the most frequent of all mutation types (Fondi et al., 2009; Zhang, 2003; Reams and Roth, 2015; Copley, 2020). The most important contribution of gene duplication is the provision of new genetic material, allowing the formation of new genes from pre-existing ones (Fani and Fondi, 2009).

Another major mechanism of gene evolution is the fusion of independent cistrons leading to bi- or multi-functional proteins (Brilli and Fani, 2004). Gene fusions characteristically involve proteins that function in a concerted manner, such as i) successive enzymes in metabolic pathways, ii) catalytic domains of enzymes and those involved in their activity regulation, and iii) DNA-binding domains and ligand-binding ones in the case, for example, of prokaryotic transcriptional regulators (Yanai et al., 2002).

A combination of gene duplication and gene fusion is gene elongation (Fig. 1), which consists of a tandem gene duplication producing two (or more) copies of the same gene, followed by the loss of the intergenic region and the mutation of the stop codon of the first copy into a sense codon, resulting in the elongation, by fusion, of the initial gene and its duplicated copy (Del Duca et al., 2020) and references therein). The newly formed gene is constituted by two paralogous modules, which might independently undergo different mutations and further duplications (Fani and Fondi, 2009) that, over time, might cover up the traces of the common origin of the two modules if the duplication events occurred very early in the evolution.

Many proteins present internal repeats of amino acid sequences, which often correspond to functional or structural domains. Thus, a gene elongation event should be considered as an evolutionary advantage, since this kind of events has occurred in many cases, and it might have shaped the structures of many genes during the first steps of molecular and cellular evolution.

The most documented and one of the most ancient cases of gene elongation involves the histidine biosynthetic genes *hisA* and *hisF*, which code for two homologous $(\beta/\alpha)_8$-barrels (also known as TIM barrels); these proteins are characterized by eight repetitions of (β-strand)-loop-(α-helix) units. The *hisA* and *hisF* genes are paralogous and may have originated from the duplication (and the subsequent divergence and fusion of the two resulting copies) of an ancestral gene, half the size of the current genes (and thus encoding a $(\beta/\alpha)_4$-half barrel) (Fani et al., 1994). These genes share the same internal organization in all histidine-synthesizing organisms, suggesting the antiquity of this elongation event and that it occurred before the appearance of the Last Universal Common Ancestor (LUCA) (Fani et al., 2007). Moreover, it was recently suggested that the two extant genes could be the result of a "cascade" of gene elongation/domain shuffling events starting from an ancestor gene coding for just one (β/α) module (Del Duca et al., 2020). According to this idea, that module might have been able to aggregate in a homo-octamer to form an unstable and inefficient complete TIM barrel. Subsequently, a cascade of three gene elongation events would have given rise to the complete ancestor of the extant TIM barrel coding genes. In addition to its evolutionary history, the *hisF* gene results particularly

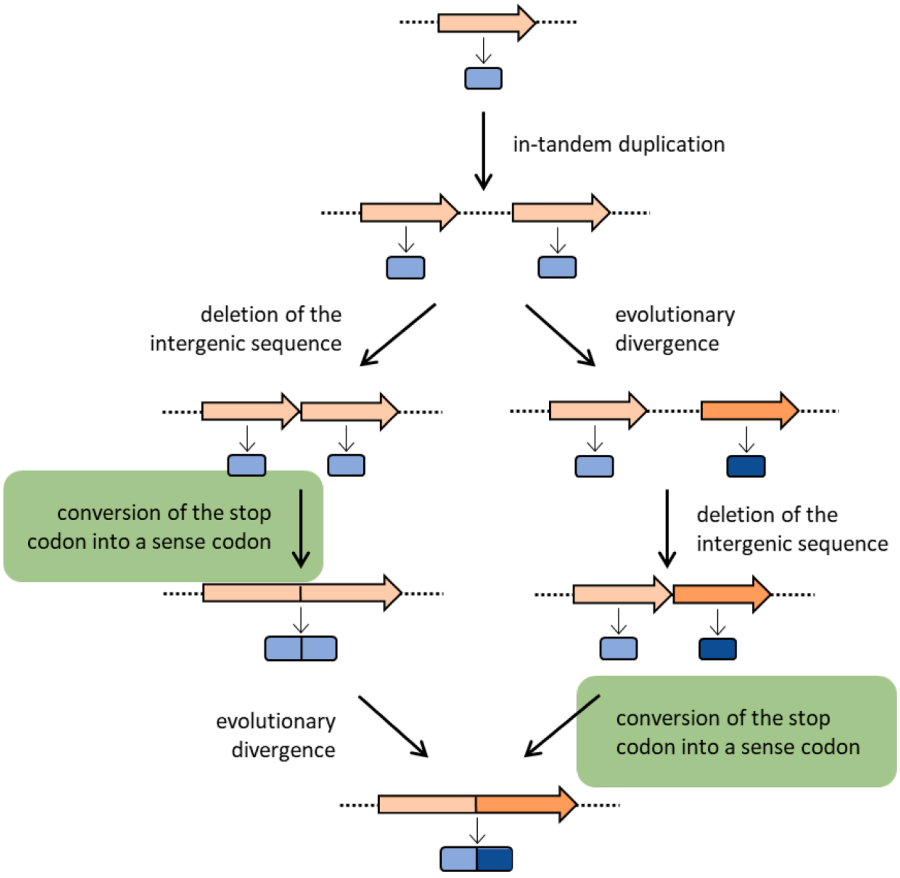

Fig. 1. Two possible evolutionary routes leading to gene elongation. Genes are represented with arrows, and the encoded proteins with rounded rectangles. This work was focused on the last step of gene elongation event (i.e., the conversion of the stop codon into a sense one), highlighted in green (). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
adapted from Del Duca et al., 2020

interesting also from a metabolic viewpoint; its product, together with the product of *hisH*, has a role in the interconnection of histidine biosynthesis with the *de novo* synthesis of purines and nitrogen metabolism, and possibly other cellular processes, such as division control, virulence, and nodule development in different bacteria (Chioccioli et al., 2020; Chioccioli et al., 2020).

Despite lots of cases regarding genes with internal sequence repetitions are reported in the literature, the gene elongation is a poorly studied evolutionary molecular mechanism. Thus, the aim of this work was to experimentally simulate the possible last step of gene elongation event occurred during *hisF* gene evolution, as shown in Fig. 1. In particular, the possible scenario after the duplication of a module half the size of the current *hisF* gene, before the fusion of the duplicated modules (i.e., the conversion of the stop codon of the first module into a sense one generating an elongated gene), was simulated by adopting selective pressure conditions. We also checked i) the time scale of elongated genes appearance, ii) the effect of different selective pressures, iii) and the type of mutation(s) responsible for gene elongation.

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   5         15         25         35         45         55         65         75
hisFab1    GGATCCATGCTCAAGA TGCGCGTCAT CCCCTGCCTG GACGTCAAGG ACGGGCGGGT CGTCAAGGGG GTCAACTTCG TCGATCTGAT
hisFab2    GGATCCATGCTCAAGA TGCGCGTCAT CCCCTGCCTG GACGTCAAGG ACGGGCGGGT CGTCAAGGGG GTCAACTTCG TCGATCTGAT
hisFab4    GGATCCATGCTCAAGA TGCGCGTCAT CCCCTGCCTG GACGTCAAGG ACGGGCGGGT CGTCAAGGGG GTCAACTTCG TCGATCTGAT

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   85         95        105        115        125        135        145        155
hisFab1         CGACGCCGGC GACCCCGTGG AGCAGGCCCG CGTCTATGAC CGGGAGGGGG CGGACGAGCT GACCTTCCTC GACATCACGG
hisFab2         CGACGCCGGC GACCCCGTGG AGCAGGCCCG CGTCTATGAC CGGGAGGGGG CGGACGAGCT GACCTTCCTC GACATCACGG
hisFab4         CGACGCCGGC GACCCCGTGG AGCAGGCCCG CGTCTATGAC CGGGAGGGGG CGGACGAGCT GACCTTCCTC GACATCACGG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   165        175        185        195        205        215        225        235
hisFab1         CCAGCCACGA GAACCGCGAC ACCATCTACG ACGTGGTGCG CCGCACCGCC GAGCAGGTGT TCATGCCGCT GACCGTCGGC
hisFab2         CCAGCCACGA GAACCGCGAC ACCATCTACG ACGTGGTGCG CCGCACCGCC GAGCAGGTGT TCATGCCGCT GACCGTCGGC
hisFab4         CCAGCCACGA GAACCGCGAC ACCATCTACG ACGTGGTGCG CCGCACCGCC GAGCAGGTGT TCATGCCGCT GACCGTCGGC

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   245        255        265        275        285        295        305        315
hisFab1         GGCGGCGTGC GCACGGTGGA CGACATCCGC AAGCTGCTGC TGGCCGGCGC CGACAAGGTG TCGATCAACA CCGCGGCCAT
hisFab2         GGCGGCGTGC GCACGGTGGA CGACATCCGC AAGCTGCTGC TGGCCGGCGC CGACAAGGTG TCGATCAACA CCGCGGCCAT
hisFab4         GGCGGCGTGC GCACGGTGGA CGACATCCGC AAGCTGCTGC TGGCCGGCGC CGACAAGGTG TCGATCAACA CCGCGGCCAT

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   325        335        345        355        365        375        385        395
hisFab1         CCACCGTCCG GAGTTCGTCC AAGAGGCGGC GGAGAAGTTC GGCGCCCAAT GCATCGTCGT CGCCATCGAC GCCAAGCAGG
hisFab2         CCACCGTCCG GAGTTCGTCC AAGAGGCGGC GGAGTAGTTC GGCGCCCAAT GCATCGTCGT CGCCATCGAC GCCAAGCAGG
hisFab4         CCACCGTCCG GAGTTCGTCC AAGAGGCGGC GGAGAAGTTC GGCGCCCAAT GCATCGTCGT CGCCATCGAC GCCAAGCAGG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   405        415        425        435        445        455        465        475
hisFab1         TCGAGCCGGG CCGCTGGGAG ATCTTCACCC ACGGCGGGCG CAAGGCCACG GGCATCGACG CCATCGAGTG GGCCAAGCGC
hisFab2         TCGAGCCGGG CCGCTGGGAG ATCTTCACCC ACGGCGGGCG CAAGGCCACG GGCATCGACG CCATCGAGTG GGCCAAGCGC
hisFab4         TCGAGCCGGG CCGCTGGGAG ATCTTCACCC ACGGCGGGCG CAAGGCCACG GGCATCGACG CCATCGAGTG GGCCAAGCGC

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   485        495        505        515        525        535        545        555
hisFab1         ATGGAGTCCT ACGGCGCCGG CGAAATCCTG CTGACCTCGA TGGACCGCGA CGGCACCAAG AGCGGCTTCG ACCTCGCCTT
hisFab2         ATGGAGTCCT ACGGCGCCGG CGAAATCCTG CTGACCTCGA TGGACCGCGA CGGCACCAAG AGCGGCTTCG ACCTCGCCTT
hisFab4         ATGGAGTCCT ACGGCGCCGG CGAAATCCTG CTGACCTCGA TGGACCGCGA CGGCACCAAG AGCGGCTTCG ACCTCGCCTT

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   565        575        585        595        605        615        625        635
hisFab1         GACCCGCAAG GTGGCCGACG GGCTGCGCAT TCCGGTCATC GCGTCGGGCG GCGTCGGCAC GCTGGACCAT CTGGTGGAAG
hisFab2         GACCCGCAAG GTGGCCGACG GGCTGCGCAT TCCGGTCATC GCGTCGGGCG GCGTCGGCAC GCTGGACCAT CTGGTGGAAG
hisFab4         GACCCGCAAG GTGGCCGACG GGCTGCGCAT TCCGGTCATC GCGTCGGGCT GAGTCGGCAC GCTGGACCAT CTGGTGGAAG

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    645        655        665        675        685        695        705        715
hisFab1          GCATCCGCGA GGGCCACGCC ACGGCGGTGC TGGCGGCCTC CATCTTCCAT TTCGGCACCT ACACCATCGG GCAGGCCAAG
hisFab2          GCATCCGCGA GGGCCACGCC ACGGCGGTGC TGGCGGCCTC CATCTTCCAT TTCGGCACCT ACACCATCGG GCAGGCCAAG
hisFab4          GCATCCGCGA GGGCCACGCC ACGGCGGTGC TGGCGGCCTC CATCTTCCAT TTCGGCACCT ACACCATCGG GCAGGCCAAG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|.
                   725        735        745        755        765        775        785
hisFab1         GCGGCGCTGG CCGAGGCCGG CATTCCGGTG CGCCCGGCCC GCATGGCGGA GGCCGCACAT GGCTGAAAGCTT
hisFab2         GCGGCGCTGG CCGAGGCCGG CATTCCGGTG CGCCCGGCCC GCATGGCGGA GGCCGCACAT GGCTGAAAGCTT
hisFab4         GCGGCGCTGG CCGAGGCCGG CATTCCGGTG CGCCCGGCCC GCATGGCGGA GGCCGCACAT GGCTGAAAGCTT
```

Fig. 2. Nucleotide sequences of the 786 bp $hisF_{ab1}$, $hisF_{ab2}$ and $hisF_{ab4}$ genes. The *Bam*HI and *Hind*III restriction sites, located respectively at the beginning and at the end of the genes, are reported in blue and were used for cloning in pTrc99A. The mismatch sites among the three genes are highlighted in grey, and the different bases from $hisF_{ab1}$ are reported in red.

2. Materials and Methods

2.1. Bacterial strains and culture conditions

The *Escherichia coli* FB8 (wild-type *E. coli* K-12 UTH1038) (Kasai, 1974) and FB182 (*hisF892*) (Goldschmidt et al., 1970) strains were used in this work. *E. coli* FB182 carries a single nucleotide deletion in position 719 of the *hisF* gene, causing a frameshift and the formation of a stop codon resulting in a shorter (243 aa vs. 258 aa of the wild-type *E. coli* HisF protein) and non-functioning enzyme (Romeo et al., 2023).

The cloning strategies were carried out with *E. coli* DH5α (F– *endA1 glnV44 thi1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM*15 Δ*(lacZYA-argF)U169 hsdR17(rK–mK +)* λ-) (laboratory stock). Cells were grown in LB (Sambrook et al., 1989) medium, supplemented with agar (1.6% w/v) and ampicillin (100 μg/mL) when required.

E. coli FB182 complementation assays and selective pressure experiments were performed on minimal medium Davis (MMD) (Davis and Mingioli, 1950) ($(NH_4)_2SO_4$ 1 g/L; K_2HPO_4 7 g/L; KH_2PO_4 2 g/L; Na_3-citrate·$2H_2O$ 0.5 g/L; $MgSO_4·7H_2O$ 0.1 g/L; pH 7.2) with agar (1.6% w/v), glucose (1% w/v), ampicillin (100 μg/mL), IPTG (50 μg/mL), and histidine (0.3–1-25 μg/mL) when required. All strains were cultivated at 37 °C.

2.2. Plasmids, genes and cloning procedures

The *hisF* gene of *Azospirillum brasilense* (NCBI ID: 56449788) (named in this work $hisF_{ab1}$, Fig. 2) was chosen for the present work. The $hisF_{ab2}$ gene (Fig. 2) was obtained from $hisF_{ab1}$ through the substitution of a single nucleotide between the two halves of the gene (355A > T), generating a stop codon. The $hisF_{ab4}$ gene (Fig. 2) was obtained from $hisF_{ab1}$ through the substitution of two nucleotides (610G > T and 612C > A) in the HisF phosphate binding domain, generating a stop codon.

All the genes were synthesized by the external company Bio-Fab Research s.r.l. (Rome, Italy). Subsequently, they were cloned between the *Bam*HI and *Hind*III restriction sites of the pTrc99A expression vector (GenBank: U13872.1). The obtained recombinant plasmids will be called hereinafter pTrc99A-$hisF_{ab1}$, pTrc99A-$hisF_{ab2}$, and pTrc99A-$hisF_{ab4}$, respectively.

E. coli DH5α chemically competent cells and *E. coli* FB182 electrocompetent cells were transformed with the ligase mixtures and recombinant plasmids. Plasmid extractions were performed using the NucleoSpin Plasmid kit (Macherey-Nagel).

2.3. Selective pressure experiments and screening of de novo mutations

E. coli FB182(pTrc99A) (as control), *E. coli* FB182(pTrc99A-$hisF_{ab2}$), and *E. coli* FB182(pTrc99A-$hisF_{ab4}$) liquid cultures were grown overnight in MMD supplemented with glucose (1% w/v), ampicillin (100 μg/mL), and histidine (25 μg/mL). OD_{600} of the cultures was measured, 1 mL of cells was washed twice with saline solution (NaCl 0.9% w/v), and resuspended in the appropriate volume of saline solution to have OD_{600} = 1. Then, 100 μL of 10^{-6} dilutions were plated on MMD plates supplemented with glucose (1% w/v), ampicillin (100 μg/mL), and histidine (25 μg/mL) to determine the vital titre. In the case of the plates under selective pressure, 100 μL of the non-diluted cell suspensions were plated on MMD plates with glucose (1% w/v), ampicillin (100 μg/mL), in absence or presence of low concentrations of histidine (0.3–1 μg/mL), and in absence or presence of IPTG (50 μg/mL). These selective pressure plates were then incubated at 37 °C and checked for 25 days, and every day the His^+ colonies were counted and isolated.

Additionally, to evaluate the effect of the growth phase on the appearance of prototrophic colonies, *E. coli* FB182(pTrc99A) and *E. coli* FB182(pTrc99A-$hisF_{ab2}$) liquid cultures were grown overnight in MMD supplemented with glucose (1% w/v), ampicillin (100 μg/mL), and histidine 25 (μg/mL). Then, OD_{600} of the two cultures was measured, and new liquid cultures starting from a OD_{600} = 0.1 were set up in MMD with glucose (1% w/v), ampicillin (100 μg/mL), and histidine (25 μg/mL). OD_{600} was measured every 30 min and samples were collected at three different growth phases: i) in the middle and ii) the end of the exponential phase, and iii) in the stationary phase. Cells were then washed and plated under selective pressure as in the previous experiment.

2.4. Screening of de novo mutations via PCR and Sanger sequencing

The amplification of the plasmid insert was performed in a 20 μL-reaction volume with 0.2 μM of primers pTrc_FOR (5′-CATCCGGCTCGTATAATG-3′) and pTrc_REV (5′-CGCTTCTGCGTTCTGATT-3′) (these primers anneal next to the multiple cloning site of pTrc99A), 1 U of DreamTaq DNA Polymerase (ThermoFisher Scientific), and 1 μL of cell thermal lysate as template. The PCR cycling was set up using an annealing temperature of 54 °C. Amplicons were visualized through a 0.8% w/v agarose gel electrophoresis.

PCR products were purified using ExoSAP-IT™ Express PCR Product Cleanup (Applied Biosystems). The sequencing reaction was performed in a 10 μL-volume with 1 μL BigDye™ Terminator v3.1 Ready Reaction Mix (Applied Biosystems), 0.32 μM of primer pTrc_FOR or pTrc_REV, and 1 μL of purified PCR product as template. Sequencing reactions were purified using BigDye Xterminator™ (Applied Biosystems), and capillary electrophoresis was run in a SeqStudio Genetic Analyzer (ThermoFisher Scientific). Sequencing data were analyzed using BioEdit (Hall, 1999).

2.5. 3D structure prediction

All the 3D protein structures reported in the present work were predicted using Phyre2 software (Kelley et al., 2015), and were visualized and modified with EzMol (Reynolds et al., 2018).

2.6. Mutants growth curves and growth rates analysis

Growth curves of His^+ revertants were performed in 96-well plates. To this purpose cells were grown overnight at 37 °C under shaking (150 rpm) in LB, supplemented with ampicillin (100 μg/mL). Then, 1 mL of culture was washed twice with saline solution (NaCl 0.9% w/v) and diluted in the appropriate volume of MMD with glucose (1% w/v), ampicillin (100 μg/mL), IPTG (50 μg/mL) and presence/absence of histidine (25 μg/mL), to reach a OD_{600} = 0.01. Cells were incubated at 37 °C and OD_{600} was measured every hour for 48 h using the Infinite M Nano (Tecan, Männedorf, Switzerland) microplates reader. Each curve was performed in triplicate.

Instantaneous growth rate of a bacterial culture was calculated according to Widdel (Widdel, 2007). Maximum growth rate (μ_{MAX}) was chosen to compare a specific mutant growth in absence and in presence of histidine in the culture medium. Since every growth curve was performed in triplicate, the average μ_{MAX} ($\overline{\mu}_{MAX}$) and the relative standard deviation were calculated. Then, the ratio between the $\overline{\mu}_{MAX}$ in absence and presence of histidine was obtained.

3. Results

3.1. Overall experimental strategy

The overall experimental strategy used in this work is schematically represented in Fig. 3. Accordingly, a plasmid expression vector containing a non-functional *A. brasilense hisF* gene (due to the presence of a stop codon in the middle of the gene) was transferred *via* transformation in an *E. coli* histidine auxotrophic strain (i.e., FB182 *hisF892*). The recombinant plasmid vector, initially unable to complement the *hisF* mutation, might restore the prototrophy of the host cell if, under selective pressure (i.e., the absence or very low concentration of histidine in the growth medium), a mutation falling in the stop codon converts it

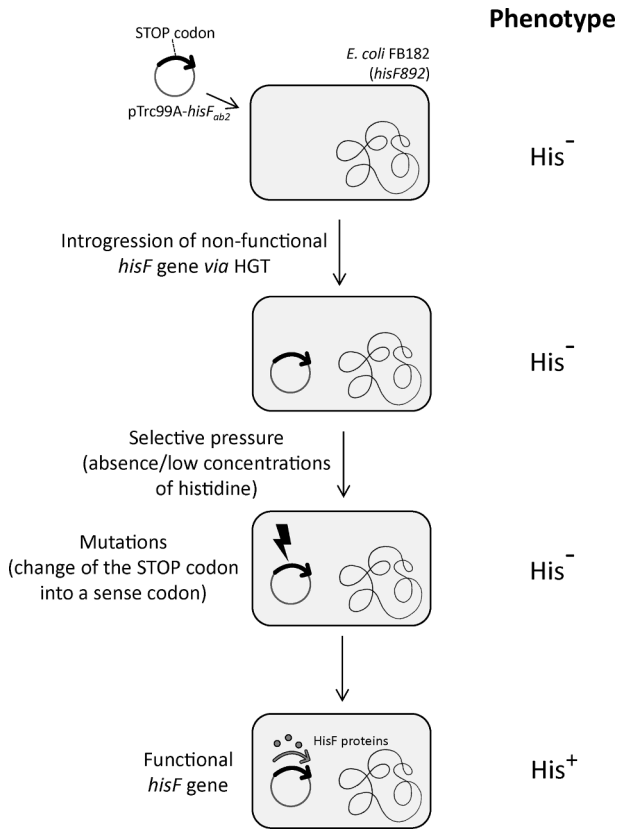

Fig. 3. Scheme of the experimental strategy.

into a sense one, thus allowing the biosynthesis of a functional HisF protein. The *hisF* gene of *A. brasilense* (referred to as *hisF*$_{ab1}$) was chosen for the present work with the aim of reducing the probability of recombination between the plasmid and the *hisF* chromosomal genes as it encodes a HisF protein that shares a degree of similarity of 59% sequence with the *E. coli* orthologue, but has a different GC content than the *E. coli hisF* gene (69.2% for *A. brasilense* and 51.9% for *E. coli*) (Fani et al., 1998).

3.2. Complementation of E. coli FB182 hisF mutation with A. brasilense hisF$_{ab1}$

The ability of *A. brasilense his* genes to complement *E. coli* histidine auxotrophy was previously demonstrated in the work of Dabizzi et al. (Dabizzi et al., 2001). However, in order to correctly set up the experimental system, we firstly checked whether the *hisF*$_{ab1}$ gene cloned into the expression vector pTrc99A was able to complement *E. coli* FB182 mutation. To this purpose, *E. coli* FB182(pTrc99A) and *E. coli* FB182 (pTrc99A-*hisF*$_{ab1}$) were streaked on MMD plates supplemented with glucose, ampicillin, in the presence or the absence of histidine, and IPTG whose presence should induce the expression of the cloned genes, since the pTrc99A vector harbors a *trc* promoter. Accordingly, data obtained after 24 h of incubation at 37 °C revealed that: i) in presence of histidine (either in presence or absence of IPTG) both *E. coli* FB182(pTrc99A) and *E. coli* FB182(pTrc99A-*hisF*$_{ab1}$) grew; i) in absence of histidine and in presence of IPTG, only *E. coli* FB182(pTrc99A-*hisF*$_{ab1}$) was able to grow; iii) in absence of both histidine and IPTG none of the two strains was able to grow.

The ability of *E. coli* FB182(pTrc99A-*hisF*$_{ab1}$) to grow in absence of histidine and in presence of IPTG confirmed the ability of *hisF*$_{ab1}$ to complement the *E. coli* FB182 *hisF* mutation and that the experimental system was correctly set up.

3.3. hisF$_{ab2}$ gene elongation study through selective pressure experiments

The *hisF*$_{ab2}$ gene differs from the wild-type *A. brasilense hisF* gene (i.e., *hisF*$_{ab1}$) by a single nucleotide substitution in position 355, in which the first adenine of the AAG codon (wild-type) was replaced with a thymine, generating the TAG stop codon. The position of the substitution was chosen on the basis of the alignment between the two halves of the *A. brasilense* HisF protein: the alignment in which the two halves showed the highest degree of similarity was selected as representative of the ancient gene elongation event involving one module half the size of the present gene (similarity between the two HisF halves of 32%). The recombinant plasmid pTrc99A-*hisF*$_{ab2}$ was then introduced by transformation into *E. coli* FB182 competent cells; transformants selected in the presence of ampicillin were then streaked on MMD plates supplemented with glucose, ampicillin, in the absence or in the presence of histidine and IPTG. After 24 h of incubation at 37 °C, *E. coli* FB182 (pTrc99A-*hisF*$_{ab2}$) did not grow in absence of histidine, even in presence of IPTG, demonstrating, as expected, the inability of *hisF*$_{ab2}$ to complement *E. coli* FB182 *hisF* mutation.

Then, two different selective pressure experiments were performed. In the first experiment (Experiment I), three replicates of *E. coli* FB182 (pTrc99A) (as control) and *E. coli* FB182(pTrc99A-*hisF*$_{ab2}$) (replicates 1, 2, and 3) liquid cultures were grown overnight and, then, each replicate was washed and plated under selective pressure, i.e. in the absence of histidine or in the presence of two different concentrations of the amino acid (0.3 and 1 μg/mL) and in the presence or absence of IPTG. For approximately 25 days, the appearance of prototrophic colonies on the different plates was checked.

In the second experiment (Experiment II), restoration of prototrophy through mutation of *hisF*$_{ab2}$ was studied in relation to the cells growth phase. *E. coli* FB182(pTrc99A) and *E. coli* FB182(pTrc99A-*hisF*$_{ab2}$) liquid cultures were set up as described in Materials and Methods, and samples were collected in correspondence of three different growth phases: in the middle (A) and at the end (B) of the exponential phase, and in the stationary phase (C) (Fig. 4). Cells were then washed and plated under selective pressure at different concentrations of histidine (0–0.3–1 μg/mL) and in the presence or absence of IPTG. The appearance of prototrophic colonies on the different plates was checked for approximately 25 days.

Data obtained are shown in Table 1 whose analysis revealed that:

i. At the end of the experiments a total of 20 His$^+$ revertants for Experiment I and 31 His$^+$ revertants for Experiment II, very likely due to mutations in the plasmid molecules, were collected (the few possible chromosomal revertants, identified by their ability to equally grow in the absence of histidine and in the presence/absence of IPTG were discarded for subsequent analyses).
ii. His$^+$ revertants appeared starting from the 4th day of incubation under selective pressure, and they continued to appear until the 16th day.
iii. Only one His$^+$ revertant out of 51 grew in the absence of histidine on MMD plates, while 13 and 37 in the presence of 0.3 μg/mL and 1 μg/mL histidine, respectively. This suggested that the number of revertants increases with the lowering of selective pressure (i.e., the increase in the concentration of histidine in the culture medium).
iv. Mutation frequency ranged between 10^{-8} and 10^{-9}, and no significant difference on the mutation frequency was observed among cells taken in the different growth phases.

3.4. Nucleotide sequencing of His$^+$ revertants

The nucleotide sequence of the insert in the *E. coli* FB182(pTrc99A-*hisF*$_{ab2}$) revertants was then determined through Sanger sequencing, as described in Materials and Methods. Data obtained with Experiments I and II are reported in Table 1 and summarized in Table 2 and Fig. 5.

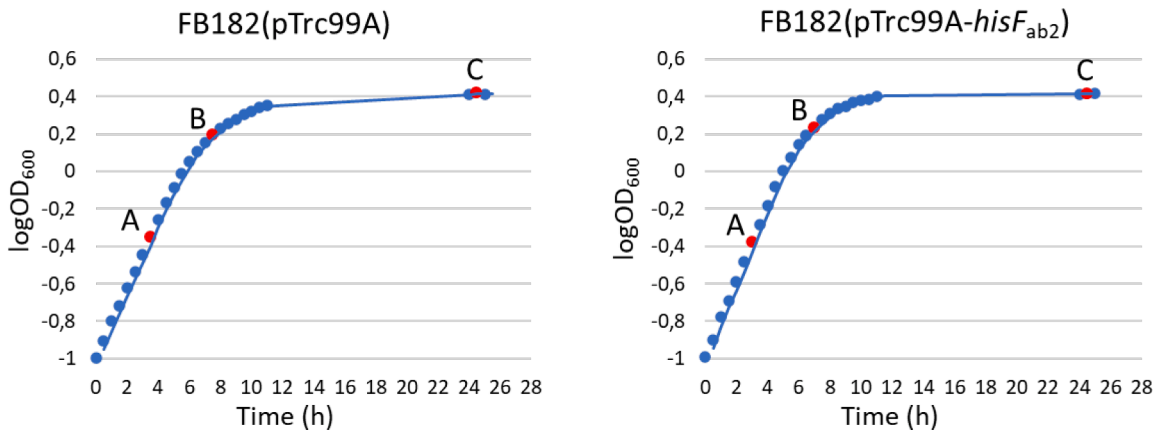

Fig. 4. Growth curves of *E. coli* FB182(pTrc99A) and *E. coli* FB182(pTrc99A-*hisF*$_{ab2}$) liquid cultures. The collection points, corresponding to the different growth phases (middle of the exponential phase - A, end of the exponential phase - B, and stationary phase - C) are indicated in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

From the obtained data it was possible to observe that:

i. All the obtained plasmid mutants gained prototrophy through a single nucleotide substitution in the stop codon located between the two *A. brasilense hisF* halves.
ii. None of the plasmid mutants restored the wild-type codon.
iii. The most frequent codon found was TTG (retrieved in about 51% of the plasmid mutants), coding for a leucine.
iv. Among all the 51 single nucleotide mutations, only 6 were transitions (those leading to the CAG codon, starting from the TAG stop codon) while the others were all transversions;
v. The first mutants to appear were those carrying the TTG codon, and to a lesser extent those with the CAG codon. Then, from the 7th day of incubation, mutants carrying the other codons started to emerge, together with the two previous ones.

3.5. Complementation of E. coli FB182 hisF mutation with recombinant plasmids obtained from His$^+$ revertants

To test whether the plasmid mutation was sufficient to render the *E. coli* FB182 cells prototrophic, plasmids were extracted from six randomly chosen mutants and used for new *E. coli* FB182 transformation. The ability of these cells to grow in absence of histidine thanks to the newly acquired plasmid was tested plating the cell cultures on MMD plates supplemented with glucose, ampicillin, in presence and absence of histidine and IPTG. Their capacity to grow in absence of histidine, but in presence of IPTG, suggested that the plasmid mutation alone was able to confer prototrophy to *E. coli* FB182.

3.6. Analysis of the possible relation between the mutants frequency and the different codon

Assuming that the differential appearance of the plasmid mutants on the basis of the codon could be due to the *E. coli* availability of charged tRNA, *E. coli* codon usage was evaluated using the Codon Usage Database (Nakamura et al., 2000); however, no correlation was observed between *E. coli* codon usage and the frequency of the different plasmid mutants (not shown).

Thus, it was hypothesized that the differential appearance of the mutants could be determined by the different growth speed, related to a better/lesser functioning HisF enzyme due to the different amino acid residue incorporated. The analysis of the predicted three-dimensional structure of the different HisF proteins shown in Fig. 6 revealed that all the plasmid mutants restored the $(\beta/\alpha)_8$-barrel structure, with a slightly different conformation of the outer protein surface that might affect the catalytic activity of the HisF enzyme.

To evaluate if the different incorporated amino acid, and the corresponding three-dimensional conformation of the HisF protein, was correlated to a different growth ability, growth curves of *E. coli* FB182 (pTrc99A-*hisF*$_{ab1}$) as positive control, *E. coli* FB182(pTrc99A) and *E. coli* FB182(pTrc99A-*hisF*$_{ab2}$) as negative controls, and all the His$^+$ revertants were performed in microplates. The ratio (r) between the average maximum growth rate ($\overline{\mu}_{MAX}$) in absence and presence of histidine, and the relative standard deviation, were calculated as reported in Materials and Methods. Data obtained are reported in Fig. 7.

As shown in Fig. 7, there is not a clear difference in the behavior of the mutants on the basis of their amino acid sequence. Nevertheless, it was possible to observe that, in general, mutants showed a reduced r value when compared to *E. coli* FB182 carrying the *A. brasilense* wild-type *hisF* gene, meaning that their ability to grow in absence of histidine is slightly reduced. However, there were some exceptions, belonging to the TTG, TCG, and TAT groups: indeed, there was a high variability among the members of these groups, suggesting that the different amino acid is not related to a different growth ability. This finding is quite surprising; we cannot *a priori* exclude the possibility that the variability in ratios between the average maximum growth rate in absence and presence of histidine among members of the TTG, TCG and TAT groups might be due to (an)other mutation(s) falling in other regions of the plasmid, such as i) in the *trc* promoter and/or ii) in the region flanking the origin of replication; such mutations could potentially affect, for example, i) the transcription efficiency and ii) the molecular mechanisms controlling the plasmid copy-number, respectively. However, even though it is quite interesting, this issue is beyond the scope of this work.

Since none of the plasmid mutants restored the wild-type codon, the conservation of the *A. brasilense hisF* AAG codon (position 355–357) was evaluated in some *A. brasilense*-related organisms. The *A. brasilense hisF* gene was used as query for a BLASTn (Camacho et al., 2009) search, and the first 100 hits were downloaded and aligned using BioEdit (Hall, 1999) through the ClustalW tool (Thompson et al., 1994) (Supplementary File S1). The AAG codon, encoding a lysine, resulted to be moderately preserved (35%), often substituted by ATG (methionine, 53%), ACG (threonine, 7%), or AAA (lysine, 5%), all coding for amino acids with different chemical properties.

3.7. hisF$_{ab4}$ retro-mutation through selective pressure experiments

The same experiment performed on *E. coli* FB182(pTrc99A-*hisF*$_{ab2}$) was repeated using the *hisF*$_{ab4}$ gene. This gene was obtained from *hisF*$_{ab1}$ through the substitution of two nucleotides generating a stop codon. These mutations were located in the sequence encoding the phosphate-binding site, the most conserved region in many $(\beta/\alpha)_8$-barrel proteins

Table 1
E. coli FB182(pTrc99A-$hisF_{ab2}$) His^+ revertants obtained in Experiment I and II. Codon replacing the original TAG stop codon, concentration of histidine, and day of appearance under selective pressure are reported. Mutated nucleotides are reported in uppercase. Mutation frequency was calculated on the basis of the vital titre and the number of obtained mutants.

Experiment	Replicate/ growth phase	Colony	Codon	Day of appearance	Selective pressure [HIS] (μg/mL)	Total n. of mutants	Initial vital titre	Mutant frequency
I	1	D2	**tTg**	5	1	7	3.80×10^{8}	6.14×10^{-8}
		D4	**Cag**	6	1			
		D5	**taT**	7	1			
		D6	**taT**	7	0			
		D8	**tCg**	8	1			
		D9	**tTg**	9	1			
		D10	**taT**	11	1			
	2	E13	**tTg**	6	1	5	3.90×10^{8}	4.27×10^{-8}
		E15	**Cag**	6	0.3			
		E16	**tCg**	9	1			
		E17	**tTg**	10	1			
		E18	**taT**	11	0.3			
	3	F6	**tTg**	4	0.3	8	4.80×10^{8}	5.56×10^{-8}
		F7	**tTg**	4	0.3			
		F8	**tTg**	5	1			
		F10	**tTg**	5	1			
		F12	**tCg**	6	1			
		F13	**tTg**	6	0.3			
		F14	**Gag**	8	1			
		F15	**tCg**	8	1			
II	A	PHA2	**tTg**	5	0.3	3	1.70×10^{8}	5.88×10^{-8}
		PHA3	**tTg**	7	1			
		PHA5	**tCg**	14	1			
	B	PHB2	**Cag**	4	0.3	17	1.77×10^{9}	3.20×10^{-8}
		PHB3	**tTg**	5	1			
		PHB4	**Cag**	5	0.3			
		PHB5	**tTg**	6	1			
		PHB6	**tTg**	6	1			
		PHB7	**tTg**	6	1			
		PHB8	**Cag**	6	1			
		PHB9	**tTg**	6	1			
		PHB11	**tTg**	8	1			
		PHB12	**tTg**	8	1			
		PHB13	**tCg**	8	1			
		PHB14	**tCg**	8	1			
		PHB15	**taC**	8	0.3			
		PHB16	**taT**	8	1			
		PHB19	**tCg**	11	1			
		PHB20	**Cag**	11	1			
		PHB21	**tCg**	16	1			
	C	PHC2	**tTg**	5	1	11	5.48×10^{9}	6.69×10^{-9}
		PHC5	**tTg**	6	0.3			
		PHC6	**tTg**	6	1			
		PHC7	**tTg**	7	1			
		PHC8	**tTg**	7	1			
		PHC9	**tTg**	7	1			
		PHC11	**tTg**	8	0.3			
		PHC12	**tCg**	9	1			
		PHC13	**tCg**	9	0.3			
		PHC14	**tTg**	10	1			
		PHC17	**tCg**	14	0.3			

Table 2
Mutants obtained with Experiments I and II. Number of mutants, codon frequency, and encoded amino acid are reported for each kind of mutation. The mutated nucleotide is reported in uppercase.

Codon	N. of mutants	Percentage of mutation (%)	Amino acid
aag (wt)			Lys
tag			Stop
tTg	26	50.98	Leu
tCg	12	23.53	Ser
Cag	6	11.76	Gln
taT	5	9.80	Tyr
taC	1	1.96	Tyr
Gag	1	1.96	Glu
Total n. of mutants	51		

(Bork et al., 1995; Del Duca et al., 2022). One of the most conserved residues of the phosphate-binding site is a glycine (position 204 in *A. brasilense* HisF) which is encoded by the codon GGC (position 610–612 in *A. brasilense hisF* gene). This codon cannot be mutated into a stop codon through a single nucleotide substitution; however, glycine is encoded also by the codon GGA, which can be mutated in the stop codon TGA. For this reason, we replaced two nucleotides (610G > T and 612C > A).

The inability of $hisF_{ab4}$ to complement *E. coli* FB182 *hisF* mutation was confirmed through the streaking of the strain *E. coli* FB182 (pTrc99A-$hisF_{ab4}$) on MMD plates supplemented with glucose, ampicillin, in presence and absence of histidine, and IPTG. After 24 h of incubation at 37 °C *E. coli* FB182(pTrc99A-$hisF_{ab4}$) did not grow in absence of histidine, even in presence of IPTG.

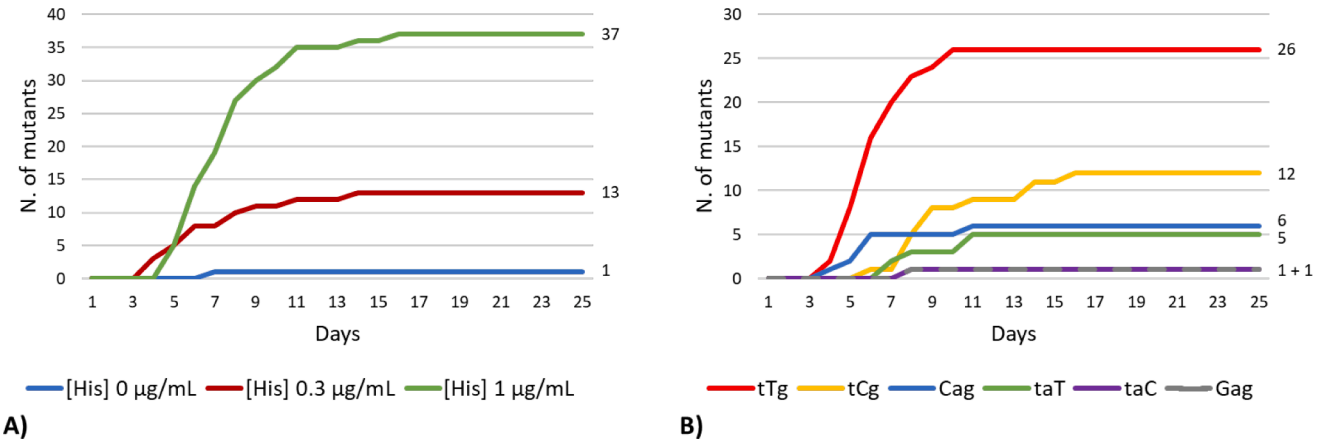

Fig. 5. (A) Timing of appearance of mutants under different selective pressures. (B) Timing of appearance of mutants on the basis of the different codon.

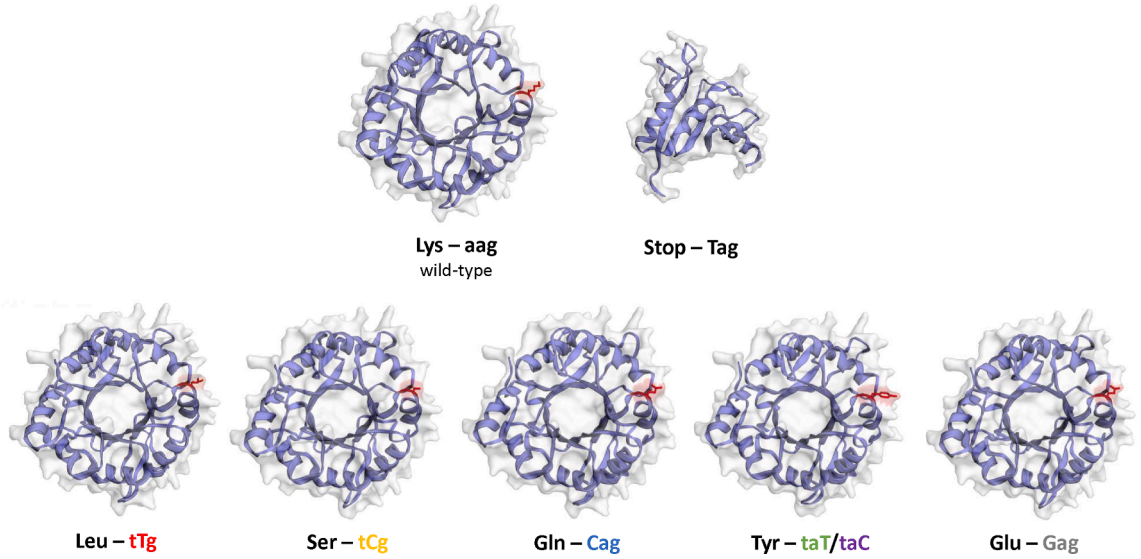

Fig. 6. Predicted three-dimensional structure of the wild type *A. brasilense* HisF protein, of the protein coded by *hisF*$_{ab2}$ gene, and of the different plasmid mutants obtained from *E. coli* FB182(pTrc99A-*hisF*$_{ab2}$) subjected to selective pressure. The amino acid that differs from the wild-type enzyme is depicted in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Then, the same experiment performed for Experiment I with *E. coli* FB182(pTrc99A-*hisF*$_{ab2}$) was performed in duplicate. The nucleotide sequence of the insert of *E. coli* FB182(pTrc99A-*hisF*$_{ab4}$) His^{+} revertants was then determined as described in Materials and Methods. The obtained plasmid mutants, the codon found in the place of the TGA stop codon, and the mutation frequency are reported in Table 3.

From the obtained data it was possible to observe that:

i. All plasmid mutants appeared between the 6th and the 8th day of incubation under selective pressure;
ii. All mutants grew on a concentration of histidine of 1 µg/mL;
iii. Mutation frequency is in the order of 10^{-9}, thus almost an order of magnitude smaller than that obtained for *E. coli* FB182(pTrc99A-*hisF*$_{ab2}$);
iv. All the obtained plasmid mutants gained prototrophy through a single nucleotide substitution in the stop codon located in the *A. brasilense hisF* phosphate-binding domain;
v. All the plasmid mutants restored the expected codon, coding for a glycine.

4. Discussion

Many present-day proteins show internal repeats of amino acid sequences, which often correspond to functional or structural domains. These internal repetitions are the outcome of one or more events of gene elongation that have occurred in many cases, and they might have shaped the structures of many genes during the first steps of molecular and cellular evolution. Indeed, it can be considered one of the main molecular mechanisms shaping the architecture of the extant genes. Gene elongation has important biological significances, which are considered evolutionary advantages. Indeed, it might allow i) the improvement of a protein function by increasing the number of active sites, ii) the acquisition of an additional function by modifying a redundant segment, obtaining a bifunctional enzyme, and/or iii) the stabilization of a protein structure, thus increasing the enzyme's catalytic activity. However, despite lots of cases of genes with internal sequence repetitions are reported in the literature, the gene elongation is a poorly studied evolutionary molecular mechanism (Del Duca et al., 2020).

The aim of the present work was to experimentally simulate the possible gene elongation event occurred during gene evolution by

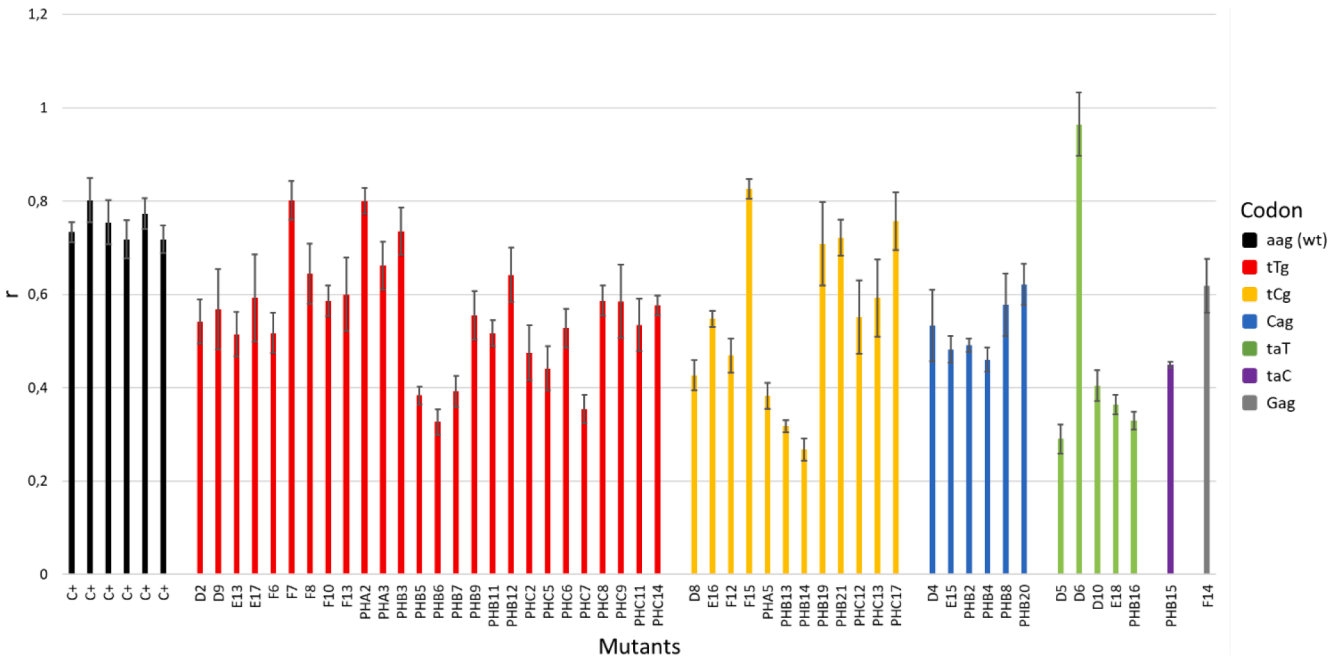

Fig. 7. Ratios (r) between the average maximum growth rate ($\bar{\mu}_{MAX}$) in absence and presence of histidine, and the relative standard deviations. In black, the r of the control strain *E. coli* FB182(pTrc99A-$hisF_{ab1}$) (one for every set up microplate). The r of the plasmid mutants is reported in red for TTG codon, yellow for TCG, blue for CAG, green for TAT, violet for TAC, and grey for GAG. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
E. coli FB182(pTrc99A-$hisF_{ab4}$) His^+ revertants, codon found in the place of the TGA stop codon, concentration of histidine, and day of appearance under selective pressure are reported. The mutated nucleotide is reported in uppercase. Mutation frequency was calculated on the basis of the number of mutants obtained for group and the vital titre.

Experiment	Replicate	Colony	Codon	Day of appearance	Histidine (μg/mL)	N. of mutants	Vital titre	Mutation frequency
I	1	D2	**Gga**	8	1	1	4.65×10^8	7.17×10^{-9}
	2						1.80×10^8	
	3	F1	**Gga**	6	1	1	3.85×10^8	8.66×10^{-9}
II	1						5.05×10^7	
	2						7.16×10^8	
	3	F13	**Gga**	8	1	1	4.00×10^8	8.33×10^{-9}

adopting selective pressure conditions and the *E. coli hisF* gene as a model system. In particular, we focused on the possible situation successive to the duplication of a module half the size of the current *hisF* gene and the evolutionary divergence of the two copies, and prior to the fusion of the two moieties.

For this reason, we firstly used the *A. brasilense hisF* gene carrying a single nucleotide mutation generating a stop codon between the two halves of the gene (named $hisF_{ab2}$). The strain *E. coli* FB182(pTrc99A-$hisF_{ab2}$) was subjected to selective pressure (i.e., low concentration/absence of histidine in the growth medium) and the obtained mutants were characterized.

Regarding the mutation frequency, it turned out to be of the order of 10^{-8}, a fairly low frequency when compared to the results obtained by Dabizzi et al. (Dabizzi et al., 2001). They studied the mutations occurred at the *A. brasilense* promoter level when inserted into histidine auxotrophic *E. coli* strains, to allow the transcription of *A. brasilense his* biosynthetic genes. In this work, the mutation frequency under selective pressure (histidine concentration of 1 μg/mL) reached values ranging between 10^{-6} and 10^{-4}. In accordance with the work of Dabizzi et al. (Dabizzi et al., 2001), the accumulation of His^+ plasmid mutants was not related to the physiological status of the cells (i.e., the growth phase), and it was strongly correlated with the strength of selective pressure (i.e., the concentration of histidine in MMD plates): the greater the selective pressure the lower the number of His^+ mutants. This finding suggested that the arisal of His^+ colonies may require a residual growth of bacteria on selective plates enabling DNA replication which, in turn, might allow the accumulation of the His^+ mutants (Dabizzi et al., 2001; Del Duca et al., 2022; Galitski and Roth, 1995).

The finding that the increase of mutants number was parallel to the decrease of the applied selective pressure might be in agreement with the early stages of molecular and cellular evolution. Indeed, if we consider the Oparin idea of the origin of life (Oparin, 1957), the first cells were heterotrophic and they thrived in the primordial soup containing abiotically synthesized nutrients, required for the survival and replication of the first living beings. Thus, it is plausible that the continuous increasing number of the early cells might have been responsible for the depletion of the available abiotically synthesized nutrients. This, in turn, should have imposed a progressively stronger selective pressure allowing the survival of those cells that acquired the ability to synthesize by themselves the compounds whose concentration was progressively decreasing in the primordial soup, rendering them less dependent on the external nutrients supply (abiotically synthesized). Hence, very likely, the early cells were not subjected to a sudden and very strong selective pressure (as it might be the complete absence of an essential nutrient), but rather to a progressive lowering of (critical) nutrients concentration, allowing the progressive appearance of mutants able to synthesize the required compounds. Since it has been proposed that the *hisA/hisF* paralogs are the results of a very ancient cascade of

elongation/duplication events (Fani et al., 1994) and that the (β/α) module might represent one of the starter types (as defined by (Lazcano and Miller, 1994), in our opinion data obtained in this work fit with the above described scenario.

The return to prototrophy was also strongly dependent on the time of incubation: the first mutants appeared after four days of incubation under selective pressure, and their number increased over time up to arrive, in some cases, to the 14th-16th day of incubation. The fact that i) no His^+ revertants appeared before four days of incubation and that ii) they keep appearing for several days, but that – once isolated – iii) they are able to normally grow in 24 h both on complete and minimal medium, suggested that these mutations were not pre-existent, but rather they originated because of the selective pressure acting in the new environment. Indeed, it was demonstrated that adaptive mutations arisen after selection had been imposed (Cairns et al., 1988; Foster, 2004), and that the presence of selective pressure increases the mutation rate per cell per day by several orders of magnitude (Alonso et al., 1999).

In all the cases obtained in the present work, the mutations occurred at the stop codon level with a single base substitution and none of the mutants restored the wild-type codon. Among 51 single nucleotide mutations, only the 11% (6/51) were transitions while the others were all transversions (88%). This is in disagreement with the knowledge that, among the base substitutions, transitions are strongly favored over transversions (96% versus 4%) (Schaaper and Dunn, 1987). However, this could be due to the fact that, in this case, the possible transitions starting from the TAG codon are CAG (found in 6 mutants out of 51), TGG (coding for tryptophan, an amino acid that carries an aromatic non polar lateral group, different from the amino acids of the other mutants) and TAA (a stop codon). However, it is still not clear why the CAG codon, despite having originated from a transition event (which should be favored), occurred only in the 11% of the cases. Moreover, it is possible to notice that all the amino acids encoded by the mutated codons (except for leucine encoded by the most frequent codon found, TTG) code for residues with polar or charged – either positively or negatively – lateral chains, suggesting that this specific region of the $(\beta/\alpha)_8$-barrel is not a conserved region and does not need particular structural constraints. This aspect was confirmed by the alignment of the *hisF* gene of some *A. brasilense* related organisms. Data obtained highlighted that the AAG codon is not highly conserved among different organisms, and that the respectively encoded amino acids show diverse chemical properties. Observing the order of mutants' appearance, it was noticed that the first appeared mutations were those leading to the TTG codon (which was also the most frequent codon observed), and the other codons started appearing from the 7th day of incubation. This led us to hypothesize a possible correlation between the codon/amino acid properties and the frequency/speed of mutants' appearance. We analyzed the possible correlations between i) the mutants' codon sequences and *E. coli* codon usage, ii) the different amino acid and the predicted three-dimensional structure of the HisF proteins, and iii) the codons and the growth ability of the strains. In all the cases the obtained results did not suggest a possible explanation for the different codon frequency/speed of mutants' appearance. Moreover, the observation that none of the plasmid mutants obtained in this work restored the wild-type codon AAG (coding for a lysine) is cryptic, even though the codon in this specific position is not highly conserved in different, albeit related, organisms. In addition, it must be considered that, even though under selective pressure adaptive mutations arise, they are mostly – if not completely – random (Maisnier-Patin and Roth, 2015), thus there is a high dose of stochasticity that does not help in identifying a basic rule in these base substitutions.

Data obtained demonstrated that at least five different amino acids occurring with a different frequency (i.e., Leu > Ser > Gln > Tyr > Glu) can restore the TIM barrel structure of *A. brasilense* HisF and its functionality, allowing host cells to grow in a minimal medium. This might imply that, during evolution, the last step of gene elongation responsible for the final "transition" to elongated genes might occur in a short time scale. Moreover, this last step might also occur with a variety of nucleotide changes, generating different codons translated into different amino acids, provided that the inserted amino acid can permit the correct TIM-barrel folding and the overall catalytic activity of the enzyme. If this is so, one should expect that a similar experiment performed with a gene harboring a stop codon instead of a sense one which is normally translated into an amino acid residue of the catalytic site should give different results (i.e., the replacing of the stop codon in a sense one coding the correct amino acid and also with a frequency lower that that obtained in the previous experiments). This aspect prompted us to replicate the experiment using a *A. brasilense hisF* gene ($hisF_{ab4}$) carrying a stop codon in a highly conserved region in many $(\beta/\alpha)_8$-barrels, the phosphate-binding site (Bork et al., 1995). The strain *E. coli* FB182 (pTrc99A-$hisF_{ab4}$) was subjected to selective pressure and the obtained results showed that all the plasmid mutants gained prototrophy through a single nucleotide substitution in the stop codon located in the phosphate-binding domain, restoring the expected codon, coding for a glycine. Thus, results allowed us to observe that, in the case of a highly conserved region, characterized very likely by biochemical constraints, the mutation is bound.

In conclusion, experiments performed in this study allowed us to mimic a possible – albeit simplified – gene elongation event occurred during the evolution of the *hisF* gene, and data obtained showed that bacterial cells are able to modify their genome in short times suggesting that genetic changes that are advantageous under selective conditions may have played a major role in metabolic evolution.

CRediT authorship contribution statement

Sara Del Duca: Formal analysis, Data curation, Investigation, Methodology, Visualization, Writing – original draft. **Alberto Vassallo:** Formal analysis, Investigation, Validation, Writing – review & editing. **Giulia Semenzato:** Investigation, Writing – review & editing. **Renato Fani:** Conceptualization, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We are grateful to Adriano Buzzati-Traverso Foundation (Pavia, Italy) that supported this work with a research fellowship grant to Sara Del Duca (grant number 841).

Appendix A. Supplementary material

Supplementary File S1: multi-alignment of the *hisF* gene of 100 *A. brasilense*-related organisms in FASTA format. Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2023.147533.

References

Alonso, A., Campanario, E., Martínez, J.L., 1999. Emergence of multidrug-resistant mutants is increased under antibiotic selective pressure in Pseudomonas aeruginosa. Microbiology 145 (10), 2857–2862. https://doi.org/10.1099/00221287-145-10-2857.

Bork, P., Gellerich, J., Groth, H., Hooft, R., Martin, F., 1995. Divergent evolution of a β/α-barrel subclass: Detection of numerous phosphate-binding sites by motif search. Protein Sci. 4, 268–274. https://doi.org/10.1002/pro.5560040213.

Brilli, M., Fani, R., 2004. The origin and evolution of eucaryal HIS7 genes: From metabolon to bifunctional proteins? Gene 339, 149–160. https://doi.org/10.1016/j.gene.2004.06.033.

Caetano-Anollés, G., Yafremava, L.S., Gee, H., Caetano-Anollés, D., Kim, H.S., Mittenthal, J.E., 2009. The origin and evolution of modern metabolism. Int. J. Biochem. Cell Biol. 41 (2), 285–297.

Cairns, J., Overbaugh, J., Miller, S., 1988. The origin of mutants. Nature 335 (6186), 142–145.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: Architecture and applications. BMC Bioinform. 10, 1–9. https://doi.org/10.1186/1471-2105-10-421.

Chioccioli, S., Del Duca, S., Vassallo, A., Castronovo, L.M., Fani, R., 2020. Exploring the role of the histidine biosynthetic hisF gene in cellular metabolism and in the evolution of (ancestral) genes: from LUCA to the extant (micro)organisms. Microbiol. Res. 240, 126555.

Chioccioli, S.; Bogani, P.; Duca, S. Del; Castronovo, L.M.; Vassallo, A.; Puglia, A.M.; Fani, R. In vivo evaluation of the interaction between the Escherichia coli IGP synthase subunits using the bacterial two-hybrid system. *FEMS Microbiol. Lett.* **2020**, *367*(14), fnaa112 , doi:10.1093/femsle/fnaa112.

Copley, S.D., 2020. Evolution of new enzymes by gene duplication and divergence. FEBS J. 287 (7), 1262–1283.

Dabizzi, S., Ammannato, S., Fani, R., 2001. Expression of horizontally transferred gene clusters: Activation by promoter-generating mutations. Res. Microbiol. 152 (6), 539–549. https://doi.org/10.1016/S0923-2508(01)01228-1.

Davis, B.D., Mingioli, E.S., 1950. Mutants of Escherichia coli requiring methionine or vitamin B12. J. Bacteriol. 60 (1), 17–28. https://doi.org/10.1128/jb.60.1.17-28.1950.

Del Duca, S., Chioccioli, S., Vassallo, A., Castronovo, L.M., Fani, R., 2020. The role of gene elongation in the evolution of histidine biosynthetic genes. Microorganisms 8 (5), 732. https://doi.org/10.3390/microorganisms8050732.

Del Duca, S., Puglia, A.M., Calderone, V., Bazzicalupo, M., Fani, R., 2022. Effect of Non-Lethal Selection on Spontaneous Revertants of Frameshift Mutations: The Escherichia coli hisF Case. Microorganisms 10, 692.

Faddetta, T., Vassallo, A., Del Duca, S., Gallo, G., Fani, R., Puglia, A.M., 2022. Unravelling the DNA sequences carried by Streptomyces coelicolor membrane vesicles. Sci. Rep. 12, 16651.

Fani, R., Brilli, M., Fondi, M., Lió, P., 2007. The role of gene fusions in the evolution of metabolic pathways: The histidine biosynthesis case. BMC Evol. Biol. 7, 1–17. https://doi.org/10.1186/1471-2148-7-S2-S4.

Fani, R., Fondi, M., 2009. Origin and evolution of metabolic pathways. Phys. Life Rev. 6 (1), 23–52.

Fani, R., Liò, P., Chiarelli, I., Bazzicalupo, M., 1994. The evolution of the histidine biosynthetic genes in prokaryotes: A common ancestor for the hisA and hisF genes. J. Mol. Evol. 38, 489–495. https://doi.org/10.1007/BF00178849.

Fani, R., Mori, E., Tamburini, E., Lazcano, A., 1998. Evolution of the structure and chromosomal distribution of histidine biosynthetic genes. Orig. Life Evol. Biosph. 28, 555–570. https://doi.org/10.1023/a:1006531526299.

Fondi, M., Emiliani, G., Fani, R., 2009. Origin and evolution of operons and metabolic pathways. Res. Microbiol. 160 (7), 502–512. https://doi.org/10.1016/j.resmic.2009.05.001.

Foster, P.L., 2004. Adaptive mutation in Escherichia coli. J. Bacteriol. 65, 21–30.

Galitski, T., Roth, J.R., 1995. Evidence that F plasmid transfer replication underlies apparent adaptive mutation. Science 268 (5209), 421–423. https://doi.org/10.1126/science.7716546.

Goldschmidt, E.P., Cater, M.S., Matney, T.S., Butler, M.A., Greene, A., 1970. Genetic analysis of the histidine operon in Escherichia coli K12. Genetics 66 (2), 219. https://doi.org/10.1093/genetics/66.2.219.

Hall, T.A., 1999. BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT. Nucleic Acids Symp. Ser. 41, 95–98.

Kasai, T., 1974. Regulation of the expression of the histidine operon in Salmonella typhimurium. Nature 249, 523–527. https://doi.org/10.1038/249523a0.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10 (6), 845–858. https://doi.org/10.1038/nprot.2015.053.

Lazcano, A., Miller, S.L., 1994. How long did it take for life to begin and evolve to cyanobacteria? J. Mol. Evol. 39, 546–554. https://doi.org/10.1007/BF00160399.

Levasseur, A., Pontarotti, P., 2011. The role of duplications in the evolution of genomes highlights the need for evolutionary-based approaches in comparative genomics. Biol. Direct 6, 1–12.

Maisnier-Patin, S., Roth, J.R., 2015. The origin of mutants under selection: How natural selection mimics mutagenesis (adaptive mutation). Cold Spring Harb. Perspect. Biol. 7 (7), a018176 https://doi.org/10.1101/cshperspect.a018176.

Nakamura, Y., Gojobori, T., Ikemura, T., 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Nucleic Acids Res. 28 https://doi.org/10.1093/NAR/28.1.292.

Oparin, A. I. The origin of life on the earth 3rd ed. *The origin of life on the earth*, **1957**.

Reams, A.B., Roth, J.R., 2015. Mechanisms of gene duplication and amplification. Cold Spring Harb. Perspect. Biol. 7 (2), a016592 https://doi.org/10.1101/cshperspect.a016592.

Reynolds, C.R., Islam, S.A., Sternberg, M.J.E., 2018. EzMol: A Web Server Wizard for the Rapid Visualization and Image Production of Protein and Nucleic Acid Structures. J. Mol. Biol. 430 (15), 2244–2248. https://doi.org/10.1016/j.jmb.2018.01.013.

Romeo, L., Esposito, A., Bernacchi, A., Colazzo, D., Vassallo, A., Zaccaroni, M., Fani, R., Del Duca, S., 2023. Application of Cloning-Free Genome Engineering to Escherichia coli. Microorganisms 11, 1–16. https://doi.org/10.3390/microorganisms11010215.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular cloning: A laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press.

Schaaper, R.M., Dunn, R.L., 1987. Spectra of spontaneous mutations in Escherichia coli strains defective in mismatch correction: the nature of in vivo DNA replication errors. Proc. Natl. Acad. Sci. U. S. A. 84 (17), 6220–6224. https://doi.org/10.1073/pnas.84.17.6220.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680. https://doi.org/10.1093/NAR/22.22.4673.

Waley, S.G., 1969. Some aspects of the evolution of metabolic pathways. Comp. Biochem. Physiol. 30, 1–11. https://doi.org/10.1016/0010-406x(69)91293-6.

Widdel, F., 2007. Theory and Measurement of Bacterial Growth A. Basic and practical aspects. Di dalam Grundpraktikum Mikrobiologie 4 (11), 1–11.

Yanai, I., Wolf, Y.I., Koonin, E.V., 2002. Evolution of gene fusions: horizontal transfer versus independent events. Genome Biol. 3 (5), 1–13. https://doi.org/10.1186/gb-2002-3-5-research0024.

Zhang, J., 2003. Evolution by gene duplication: An update. Trends Ecol. Evol. 18 (6), 292–298.