BULLETIN N° 252
ACADÉMIE EUROPÉENNE
INTERDISCIPLINAIRE
DES SCIENCES
INTERDISCIPLINARY EUROPEAN ACADEMY OF SCIENCES

Lundi 6 décembre 2021 (en format mixte présence-distance) :

A 15h 30
Conférence
« Que nous dit le récent rapport du GIEC ? »
Par le Pr François-Marie BRÉON
Chercheur et directeur adjoint au Laboratoire des Sciences du Climat et de l'Environnement
Unité mixte de recherche (UMR 8212) /CEA/CNRS/Université de Versailles St-Quentin (UVSQ)
Institut Pierre-Simon Laplace (IPSL)

Notre Prochaine séance aura lieu le lundi 10 janvier 2022 de 15h30 à 18h00
Salle Annexe Amphi Burg
Institut Curie, 13 rue Lhomond - 75005 Paris

Elle aura pour thème :

« Considérations de biologie évolutive sur l'interdépendance
et la résilience dans le monde vivant »
par le Pr. **Eric BAPTESTE**
Directeur de Recherche au CNRS
UMR 7138 – Evolution/ Adaptation, Intégration, Réticulation et Evolution
Sorbonne Université/ Campus Pierre et Marie Curie

ACADÉMIE EUROPÉENNE INTERDISCIPLINAIRE DES SCIENCES
INTERDISCIPLINARY EUROPEAN ACADEMY OF SCIENCES

Décembre 2021

N°252

TABLE DES MATIERES

Prochaine séance : lundi 10 janvier 2022 de 15h30 à 18h00

« Considérations de biologie évolutive sur l'interdépendance et la résilience dans le monde vivant »
par le Pr. **Éric BAPTESTE**
Directeur de Recherche au CNRS
UMR 7138 – Évolution/ Adaptation, Intégration, Réticulation et Évolution
Institut de Biologie Paris-Seine
Sorbonne Université/ Campus Pierre et Marie Curie

ACADÉMIE EUROPÉENNE INTERDISCIPLINAIRE DES SCIENCES
INTERDISCIPLINARY EUROPEAN ACADEMY OF SCIENCES

Séance du Lundi 6 décembre 2021 mixte présence-distance

La séance est ouverte à 15h30, **sous la Présidence de Victor MASTRANGELO**

- **étaient présents physiquement nos Collègues membres titulaires** : Gilbert BELAUBRE, Jean BERBINAU, Éric CHENIN, Françoise DUTHEIL, Michel GONDRAN, Irène HERPE-LITWIN, Marie-Françoise PASSINI, Jacques PRINTZ, Jean SCHMETS, Jean-Pierre TREUIL

- **ainsi que notre Collègue membre correspondant:** Jacky ROUSSELLE

- **et connectés à distance, nos Collègues** : Jean-Louis BOBIN, Anne BURBAN, Gilles COHEN-TANNOUDJI, Ernesto DI MAURO, Jacques FLEURET, Christian GORINI, Johanna HENRION-LATCHÉ, Abdel KENOUFI, Édith PERRIER, Dominique PRAPOTNICH, Benoist PRIEUR, Gérard VAUTRIN

Conférence du Professeur François-Marie BRÉON : « Que nous dit le récent rapport du GIEC ? »

1. Présentation du Conférencier par notre président Victor MASTRANGELO

« Le Professeur Bréon est physicien de formation, ancien élève de l'École Normale Supérieure Ulm, chercheur et directeur-adjoint au laboratoire des Sciences du Climat et de l'Environnement, Unité mixte de recherche (UMR 8212) /CEA/CNRS/Université de Versailles Saint-Quentin et l'un des laboratoires de l'Institut Pierre-Simon Laplace (IPSL)

Il est spécialiste des observations spatiales pour l'étude du climat. ~~Vous êtes~~ Il est le responsable scientifique de la mission MicroCarb qui sera lancée fin 2022. Rappelons que **L**'objectif de MicroCarb est de cartographier, à l'échelle planétaire, les sources et puits du principal gaz à effet de serre : le CO2.

Il a été auteur du cinquième rapport du GIEC (**Groupe d'experts intergouvernemental sur l'évolution du climat**, **GIEC** ; en anglais : Intergovernmental Panel on Climate Change, IPCC).

Il est auteur ou co-auteur de plus de 120 publications dans des journaux à comité de lecture.
Il est par ailleurs président de l'Association Française pour l'Information Scientifique. »

2. Conférence du Pr BRÉON

Voici le résumé en français de la conférence du Professeur François-Marie BRÉON :

QUE NOUS DIT LE RÉCENT RAPPORT DU GIEC ?

Le changement climatique est une réalité. On peut constater la multiplication des canicules, la disparition de nombreux glaciers, la hausse du niveau des mers, et le changement de fréquence de certains phénomènes météorologiques extrêmes. Les climatologues affirment que c'est là la conséquence de la hausse des gaz à effet de serre dans l'atmosphère. Cette affirmation repose sur la compréhension des mécanismes qui sont à l'œuvre. Elle est aussi confirmée par la connaissance des changements climatiques passés, qui n'ont pas la même vitesse.

Ce que on observe aujourd'hui avait été annoncé depuis plus de 20 ans sur la base des simulations climatiques. Depuis, les modèles ont été améliorés permettant de tenir compte de processus complexes qui n'étaient pas nécessairement pris en compte lorsque les ordinateurs n'avaient pas la puissance d'aujourd'hui.

Pour le futur, les simulations climatiques annoncent une augmentation de l'amplitude des modifications climatiques avec des conséquences qui peuvent être désastreuses dans de nombreuses régions du monde.

Le séminaire s'appuiera essentiellement sur les conclusions du 6ème rapport du GIEC (groupe "physique du climat") qui est sorti en Août 2021. Il tentera de montrer ce qui est bien compris, et ce qui est plus incertain pour anticiper les changements climatiques futurs.

Trois articles publiés, en particulier dans le cadre de l'Afis (Association française pour l'information scientifique) :
https://www.afis.org/Changement-climatique-l-etat-des-connaissances-scientifiques
https://www.afis.org/Interview-de-Francois-Marie-Breon-journal-l-Express
https://www.afis.org/Le-climato-denialisme-n-est-pas-mort

Le Pr BRÉON rappelle la position du GIEC (Groupe d'experts Intergouvernemental sur l'Évolution du Climat) à l'interface entre science et politique. Les rapports sont rédigés par des experts dans un esprit strictement scientifique : ce sont des synthèses élaborées à partir de la littérature scientifique. En revanche, pour chaque rapport, le « résumé à l'intention des décideurs » est négocié et approuvé par les représentants des États, il combine donc des considérations scientifiques et des considérations politiques, et il relève d'un compromis entre science et politique. Le Pr BRÉON précise aussi que les auteurs des rapports ne sont pas rémunérés par le GIEC : ils consacrent une partie de leur temps de travail à cette activité, et sont donc rémunérés par leurs organismes de rattachement (dans son cas : le CEA). Il indique que le GIEC est organisé en trois groupes : Groupe 1 : science du climat ; Groupe 2 : Impact et vulnérabilité ; Groupe 3 : Atténuation du changement climatique. François-Marie BRÉON a contribué à rédiger le 5ème rapport du Groupe 1 ; celui-ci a publié cet été son 6ème rapport, qui est l'objet central de sa conférence.

Le rapport estime une élévation de la température moyenne du globe de 1°C depuis 1850, avec une nette accélération depuis 1950. Il intègre les évolutions de plusieurs

paramètres directement mesurables : la concentration moyenne de CO_2 dans l'atmosphère, la plus élevée depuis au moins 2 millions d'années, la montée du niveau des mers, la plus rapide depuis au moins 3.000 ans, la surface de la banquise arctique, la plus réduite depuis au moins 1.000 ans, et le recul des glaciers, sans précédent depuis au moins 2.000 ans. Et l'augmentation de phénomènes observables : la fréquence et l'intensité des chaleurs extrêmes, la fréquence et l'intensité des précipitations, l'augmentation de la sécheresse dans certaines régions, la fréquence des incendies, et le réchauffement, l'acidification et la désoxygénation des océans. Globalement, les changements climatiques récents sont généralisés, rapides et s'intensifient ; ils sont sans précédent depuis des milliers d'années.

L'interprétation et l'anticipation des changements s'appuient sur des modèles alliant physique, chimie et biologie, analogues aux modèles météorologiques, mais intégrant des modules qui prennent en compte les processus lents comme l'évolution des courants marins, des calottes polaires, ou de la distribution de la végétation.

L'effet de serre est connu depuis des siècles : l'atmosphère est pratiquement transparente au rayonnement solaire ($\lambda < 3$ µm) mais largement opaque au rayonnement infrarouge ($\lambda > 4$ µm). Bien que constituant moins de 1% de l'atmosphère, ce sont les molécules de plus de 2 atomes qui causent cette opacité, par une combinaison d'absorption et de réflexion vers la Terre (H_2O, CO_2, CH_4, N_2O). Depuis les années 90, les preuves de l'influence humaine sur l'effet de serre s'accumulent, et aujourd'hui cette influence ne fait plus de doute. A noter que l'expression « effet de serre » est en fait une image : une serre empêche le refroidissement par convection ; alors que pour l'atmosphère il s'agit à proprement parler de « forçage radiatif » : un bilan positif du rayonnement reçu par rapport au rayonnement réémis. A noter également qu'il y existe aussi une influence humaine, plus faible, en faveur du refroidissement, par nos émissions d'aérosols, lesquels réfléchissent vers l'espace une partie du rayonnement solaire.

Parmi les processus à l'œuvre dans le réchauffement de l'atmosphère, il y a des rétroactions positives, comme la diminution de l'albédo avec la réduction des surfaces enneigées, et comme l'augmentation de la contenance en vapeur d'eau de l'atmosphère, qui influent toutes deux sur le forçage radiatif et renforcent le réchauffement. Cette augmentation de capacité en H_2O favorise en outre l'évaporation, et donc à la fois la fréquence des sécheresses et l'intensité des précipitations.

Le 6ème rapport du GIEC affine l'analyse, via un découpage des continents en 45 régions, où sont examinés divers paramètres observables sur lesquels le changement du climat peut influer - fréquence des canicules, précipitations extrêmes, dates de floraison des cerisiers, dates des vendanges, etc. -. L'évolution de chaque paramètre est estimée - en augmentation, en diminution ou indéterminée - ainsi que la probabilité de l'influence anthropique - forte, moyenne, faible, indéterminée -.

Les prédictions actuelles prévoient 1,5 °C d'accroissement en 2040 par rapport au début de l'ère industrielle ; rester ensuite au-dessous de 2 °C d'accroissement, comme les signataires des accords de Paris lors de la COP21 s'y sont engagés, est encore possible moyennant des efforts importants. Mais les prévisions moyennes dépassent les 2 °C dès 2050 et donnent un réchauffement tendant vers 3 °C vers 2100. Et les plus pessimistes approchent les 5 °C en 2100. Il faut noter que le réchauffement n'est pas uniforme : il est plus fort sur les terres et sur les régions polaires. Le niveau des mers, sous l'effet combiné de la dilatation des eaux et de la fonte des glaces, monte au rythme de 3 à 4 mm par an, en légère accélération, et cette hausse se poursuivra au cours des prochains siècles, jusqu'à plusieurs mètres, même si l'on stabilise le climat.

Le changement du climat impactera la santé, l'agriculture, et les écosystèmes des générations futures ; les mesures pour le contenir impactent les transports, l'agriculture et globalement l'économie des générations actuelles. Il y a donc un compromis politique nécessaire entre les impacts respectifs du changement climatique et des mesures pour le limiter, et entre les impacts sur les générations actuelles et sur les générations futures. Et ces impacts ne sont pas les mêmes entre les pays, selon leur géographie et selon leur économie, et globalement, les pays les plus en cause ne sont pas les plus impactés, ce qui rend les négociations internationales particulièrement difficiles.

Pour limiter le réchauffement, nous devons limiter nos émissions de CO_2, de CH_4, et de N_2O. Cela suppose de limiter notre utilisation des énergies fossiles, essentiellement grâce à plus de sobriété et d'efficacité, et en augmentant l'usage d'une électricité non carbonée. Il y a un compromis politique à trouver entre forte sobriété et transition massive vers l'énergie électrique, laquelle sera largement, à moyen terme, d'origine nucléaire.

L'exposé suscite de très nombreuses questions :

Jean Berbinau se rappelle que dès Juillet 2001, lors d'un exposé au Ministère des Finances, un membre du GIEC présentait déjà un réchauffement de 2 °C comme inévitable ; il se demande pourquoi un objectif de 1,5 °C a été adopté lors de la COP21, alors qu'il semble irréaliste. F.-M. BRÉON répond que cet objectif très ambitieux a peut-être été jugé utile pour mieux assurer d'atteindre un objectif moins ambitieux. Il précise ensuite que physiquement, cet objectif est accessible, mais que l'inertie de nos sociétés empêchera vraisemblablement de l'atteindre.

Eric Chenin évoque la dispersion naturelle des énergies solaire et éolienne, et la dépense d'énergie nécessaire pour lutter contre cette dispersion. Il se demande si l'on peut calculer l'énergie minimale correspondant à la lutte contre l'entropie pour concentrer ces énergies dans des machines. Par exemple, un hectare de colza absorbe de l'énergie solaire pendant une année pour produire 600 litres d'huile utilisée pour faire tourner un moteur de deux litres pendant 100 heures : peut-on calculer l'énergie minimale correspondant à la diminution d'entropie correspondant à cette concentration de l'énergie dans l'espace et dans le temps ? Est-ce qu'on pourrait ainsi, indépendamment des méthodes utilisées pour capter ces énergies dispersées et de leurs rendements propres, identifier une limite théorique à la quantité d'énergie utile récupérable à partir du soleil et du vent ? F.-M. BRÉON répond qu'a priori oui, mais que l'énergie solaire, par exemple, est disponible sur Terre dans de tels volumes que la limite ne semble pas se situer dans une limite théorique de proportion récupérable, mais dans une limite pratique liée à la quantité de matériaux nécessaires pour la capter et la transporter. Il pense que ce n'est pas la physique qui impose la limite, mais des considérations pratiques liées aux quantités de capteurs et à l'infrastructure de transport nécessaires. Françoise Dutheil indique que l'énergie solaire est disponible à hauteur d'1 KW / m^2 (à l'équateur à midi) : le problème est donc la surface disponible près des zones d'utilisation de l'énergie. F.-M. BRÉON ajoute que la difficulté dans l'exploitation de l'énergie solaire ne ressort pas de la physique, mais de considérations pratiques techniques, géographiques et géopolitiques. Pour l'éolien, F.-M. BRÉON indique que l'énergie disponible en moyenne est d'1 W / m^2 : les éoliennes ne peuvent donc pas être installées avec une trop grande densité, mais l'énergie, là aussi est disponible en quantité : 1 km^2 équipé peut en théorie fournir 1 MW. Pour l'éolien aussi, la limite n'est pas imposée par la physique, mais par des considérations pratiques techniques, géographiques et sociales associées aux capteurs et au transport. Dans les deux cas - solaire et éolien -, l'intermittence pose en outre un problème aigu de stockage.

Jacky Rousselle évoque la déforestation et se demande quel est son impact en termes d'émission de CO_2, il évoque aussi l'albédo des sols selon leurs usages (forêt, cultures, etc.) et se demande ce que les connaissances actuelles permettent d'envisager pour améliorer les choses. F.-M. BRÉON indique que les émissions de CO_2, ces dernières années, sont de l'ordre de 10% des émissions associées aux énergies fossiles. Mais la végétation terrestre reste un puits de carbone, même en tenant compte de la déforestation : sa croissance est renforcée par l'augmentation du taux de CO_2, et elle absorbe aujourd'hui environ ¼ de nos émissions (environ 10 milliards de tonnes absorbées pour 40 milliards de tonnes émises). Les océans absorbent aussi environ 10 milliards de tonnes, essentiellement par dissolution dans l'eau. Pour ce qui est de l'albédo, quand on remplace une forêt par un champ de maïs, l'albédo augmente, en faveur du refroidissement du climat, mais l'ordre de grandeur est très faible : un ordre de grandeur en dessous des phénomènes que l'on vient d'évoquer.

Jean-Pierre Treuil évoque les « climato-sceptiques » et leurs arguments. Par exemple, l'existence d'un « petit âge glaciaire », du refroidissement du climat entre l'an 1000 et 1800, ou le doute sur la précision de l'estimation de la température moyenne, de l'ordre du dixième de degré pour une température de l'ordre de 300 en degrés Kelvin. Il demande comment les scientifiques peuvent s'assurer d'une telle précision pour une température moyenne du globe, compte tenu également des variations importantes au cours de l'année et selon les endroits. F.-M. BRÉON répond que sur la France par exemple, la température moyenne annuelle varie de la même manière dans toutes les régions : il y a une corrélation spatiale de l'évolution de la température. Donc la mesure de la température n'a pas besoin d'une trop grande précision spatiale : sur les terres, une précision spatiale de l'ordre de 1.000 km suffit, et sur les océans, la corrélation spatiale est encore plus importante. Donc, des capteurs assez épars suffisent. Plusieurs groupes ont estimé la température moyenne de manière indépendante : ils évaluent leur incertitude à environ 1/10 de degré, et ils obtiennent le même résultat au 1/10 de degré près. L'incertitude sur l'estimation de la température moyenne est donc de l'ordre du 1/10 de degré Celsius. Donc, quand on observe un réchauffement de 1 °C, on est très au-dessus de l'incertitude. En 1980, quand on estimait le réchauffement à environ 2/10 de degré, on pouvait encore douter de la précision de cette estimation ; mais maintenant, le réchauffement estimé est très au-dessus de l'ordre de grandeur de l'incertitude. Quant au « petit âge glaciaire », il y a des incertitudes sur son amplitude, et on peut proposer plusieurs causes : le déplacement des courants marins au cours des siècles, et des éruptions volcaniques plus fréquentes que sur les derniers 50 ans, et peut-être un rayonnement solaire plus faible. En conclusion, il y a des phénomènes encore mal compris, il y a eu des variations du climat dans le passé, mais cela ne permet pas de réfuter le réchauffement actuel estimé, ni son imputation à l'homme.

Gilbert Belaubre évoque le niveau des mers, et demande quels sont les maximas et les minimas que la Terre a connus, et quelles sont les perspectives au point de vue de la surface des terres émergées. F.-M. BRÉON répond que le niveau est monté de 130 mètres depuis la dernière période glaciaire. En termes de perspectives, la hausse est actuellement de 3 mm par an, et une hausse de plusieurs mètres dans le long terme est inévitable. Mais la hausse ne sera pas de l'ordre de celle qui a eu lieu depuis la dernière glaciation, car le volume de glace encore mobilisable est déjà réduit. Si tout l'Antarctique fondait, le niveau monterait de 70 mètres supplémentaires, mais cela ne devrait pas se produire.

Jacques Printz questionne la corrélation entre température et CO_2. F.-M. BRÉON répond qu'il y a des équations physiques qui relient émissions et concentration de CO_2. La dérivée de la concentration est proportionnelle à la différence entre les émissions et les puits, l'incertitude venant plutôt des puits. Ensuite, la relation entre concentration de CO_2 et température relève de connaissances bien établies en physique. La physique de l'effet de serre est parfaitement

comprise : le niveau de rayonnement infrarouge de la Terre est fonction de sa température, et l'absorption-réflexion de ce rayonnement par l'atmosphère est bien compris.

Eric Chenin demande ensuite si l'on peut envisager de réaliser toutes les applications des énergies fossiles avec de l'énergie nucléaire. F.-M. BRÉON indique que l'électricité nucléaire constitue 10 % de l'électricité dans le monde, et l'électricité représente actuellement ¼ de l'énergie dans le monde, donc aujourd'hui, le nucléaire fournit 1/40 de l'énergie dans le monde. Remplacer toutes les sources d'énergie par du nucléaire n'est pas impossible physiquement, mais certaines applications sont actuellement impossibles à réaliser avec du nucléaire, par exemple le transport aérien. Sur le plan des réserves, pour le nucléaire actuel, il y a une centaine d'années de réserves d'uranium ; mais si l'on passe aux technologies dites des neutrons rapides, on multiplie cette durée par 100. F.-M. BRÉON ne pense pas que l'on remplace entièrement les énergies fossiles par du nucléaire, et cela ne lui semble pas nécessaire, grâce aux énergies renouvelables, qui vont se développer ; et puis il y aura des adaptations à faire dans notre consommation énergétique, notamment vers plus de sobriété.

Jean Berbinau revient sur les 5,5% de baisse des émissions de CO_2 dues à la pandémie de Covid-19 entre 2019 et 2020. Il se demande si cette baisse est durable. F.-M. BRÉON montre l'évolution entre 2019 et 2021 : la baisse au niveau mondial n'est que de 1%, essentiellement sur le transport aérien et le transport terrestre. Donc la baisse ne semble pas durable. En revanche, la pandémie a montré que des changements importants sont possibles dans nos comportements : confinement, télétravail, réunions à distance.

Benoit Prieur évoque le méthane océanique, non mentionné dans l'exposé. Il rappelle qu'il y a une quantité importante de biomasse océanique enfouie sous les sédiments et donc anaérobie, deux fois plus importante en masse que la biomasse aérienne et elle produit en permanence des quantités considérables de méthane, qui est parfois stocké dans des poches de méthane hydraté. Ce méthane peut parfois s'échapper, en quantité, et il provoque alors une augmentation de l'effet de serre (temporaire car le méthane ne perdure que 30 ans dans l'atmosphère, mais important car son absorption est 30 fois celle du CO_2). Il pense que l'on ne tient pas assez compte du rôle actuel et potentiel du méthane dans l'effet de serre, et des quantités de méthane produit par la biomasse aérienne et sous-terraine. F.-M. BRÉON indique que le transfert radiatif du méthane est bien quantifié et que la concentration du méthane dans l'atmosphère est connue avec précision, y compris le niveau de son augmentation depuis l'ère préindustrielle. Le méthane est cause d'environ 30% de l'effet de serre, mais pour l'avenir, il y a une incertitude sur les quantités de méthane qui pourront être relâchées.

Enfin, Jean Berbinau évoque les variations de l'angle entre l'axe de la Terre et le plan de l'écliptique. F.-M. BRÉON indique que l'angle est actuellement de 23°, mais qu'il a varié effectivement, sur des millénaires, de 18° à 25° dans le passé, et que les glaciations ont été causées par ces variations.

Documents

p.10 : le résumé de la conférence d'Eric Bapteste

p.11 : en relation avec la conférence d'Eric Bapteste, l'article « **Modeling the evolution of interconnected processes: It is the song and the singers; Tracking units of selection with interaction networks** » (Eric Bapteste et François Papale)

p.25 : en relation avec la conférence d'Eric Bapteste, l'article **« Towards a Dynamic Interaction Network of Life to unify and expand the evolutionary theory »** (Eric Bapteste et Philippe Huneman)

p.32 : l'article de notre collègue Ernesto Di Mauro : « **The role of borosilicate glass in Miller–Urey experiment »** (Joaquín Criado-Reyes, Bruno M. Bizzarri, Juan Manuel García-Ruiz, Raffaele Saladino & Ernesto Di Mauro)

Pr. Eric BAPTESTE

Directeur de Recherche au CNRS
UMR 7138 – Evolution/ Adaptation, Intégration, Réticulation et Evolution
Institut de Biologie Paris-Seine
Sorbonne Université/ Campus Pierre et Marie Curie

Résumé

Considérations de biologie évolutive sur l'interdépendance et la résilience dans le monde vivant

Dans cet exposé, je commencerai par rappeler l'ubiquité des relations de dépendances voire d'interdépendances dans le monde vivant, puisqu'il y a des interactions partout en biologie. Je considérerai que face à ce constat, le travail d'un évolutionniste pourrait être de chercher à comprendre comment les interactions ont façonné le monde vivant. Je reviendrai sur quelques découvertes à ce sujet, parcourant l'histoire des premières cellules aux organismes contemporains, évoquant le transfert latéral de gènes, la vie des communautés microbiennes et les symbioses, brossant ainsi le portrait d'un monde où les organismes s'entrecroisent et se co-construisent. Je rappellerai que ce qui est vrai à l'échelle des organismes est également vrai à une échelle plus réduite, puisque les organismes et les communautés en interactions sont-elles mêmes composées de réseaux d'interactions moléculaires dont je donnerai quelques exemples, en soulignant que la biologie des uns est parfois manipulée par celle des autres. Ces cas d'étude établiront que les interactions ont des conséquences directes sur la stabilité et le devenir d'un hôte, abordant finalement la question de la stabilité/de la dynamique/de la résilience des interactions et de leurs effets dans le monde vivant. Je proposerai une piste simple pour modéliser cette complexité au moyen de réseaux, capable d'étendre la théorie de l'évolution depuis une science de l'évolution des organismes et des espèces vers une théorie incluant jusqu'à l'évolution des écosystèmes et des communautés, avant de me questionner sur la possibilité d'étendre ce type de modèle pour penser la stabilité du réseau auquel notre espèce appartient.

Abstract

An evolutionary biology viewpoint on the interdependence and resilience in the living world

In this talk, I will recall the ubiquity of dependency relationships, hinting at some interdependencies within the living world, because interactions are everywhere in biology. I will consider, that to face this complexity, an evolutionary biologist might try to understand how interactions shaped the living world. I will present some discoveries made on that front about early and contemporary microbial evolution, mentioning horizontal gene transfer, the communal lifestyle of microbes and symbioses, to conclude that organisms' fates are intertwined and involve co-construction and sometimes manipulative processes. I will also provide examples of interactions at the molecular level. Based on these case-studies, I will argue that interaction have direct consequences on the stability of evolving entities, before turning to the issue of the stability/dynamics and resilience of these interactions and of their effects in the living world. I will propose a simple network-based modeling for this complexity, able to expand the current evolutionary theory from a science of evolving organisms and species to a science including ecosystem and communities evolution. Finally, I will question whether this gives us directions to think about the stability of the network to which our own species belongs.

PROBLEMS & PARADIGMS

Prospects & Overviews

BioEssays WILEY

Modeling the evolution of interconnected processes: It is the song and the singers

Tracking units of selection with interaction networks

Eric Bapteste[1] | François Papale[2] (ORCID)

[1] Institut de Systématique, Evolution, Biodiversité (ISYEB), Sorbonne Université, CNRS, Museum National d'Histoire Naturelle, EPHE, Université des Antilles, 7, quai Saint Bernard, Bâtiment A 4ème étage, pièce 427, Paris 75005, France

[2] Departement of Philosophy, University of Montreal, 2910 Édouard-Montpetit blvd, Montréal, QC H3C 3J7, Canada

Correspondence
Eric Bapteste, Institut de Systématique, Evolution, Biodiversité (ISYEB), Sorbonne Université, CNRS, Museum National d'Histoire Naturelle, EPHE, Université des Antilles; 7, quai Saint Bernard, Bâtiment A 4ème étage, pièce 427, 75005 PARIS, France.
Email: epbapteste@gmail.com

Abstract

Recently, Doolittle and Inkpen formulated a thought provoking theory, asserting that evolution by natural selection was responsible for the sideways evolution of two radically different kinds of selective units (also called Domains). The former entities, termed singers, correspond to the usual objects studied by evolutionary biologists (gene, genomes, individuals, species, etc.), whereas the later, termed songs, correspond to re-produced biological and ecosystemic functions, processes, information, and memes. Singers perform songs through selected patterns of interactions, meaning that a wealth of critical phenomena might receive novel evolutionary explanations. However, this theory did not provide an empirical approach to study evolution in such a broadened context. Here, we show that analyzing songs and singers, using patterns of interaction networks as a common ontology for both, offers a novel, actionable, inclusive and mathematical way to analyze not only the re-production but also the evolution and fitness of biological and ecosystemic interconnected processes.

KEYWORDS
earth system science, evolutionary biology, evolutionary systems biology, interactions networks, units of selection

INTRODUCTION

Recently, Doolittle and Inkpen introduced a thought provoking hypothesis about evolution by natural selection, entitled "It is The Song, Not The Singers" (ITSNTS).[1] This hypothesis generalizes the condition for evolution by natural selection (ENS), a major scientific concept introduced by Darwin in 1859 to explain the evolution of species.[2] ITSNTS expands this critical theoretical framework way beyond the evolution of monospecific populations to further address the evolution of processes, for instance, functions realized by communities within ecosystems. Thanks to ITSNTS, evolutionary biology can now seek to make

sense not only of the history of species, but also of the history of processes or patterns of interaction such as stable and recurrent biogeochemical cycles, for example, nitrogen fixation, operated by interactions between biotic and abiotic components.[1] The evolution of such patterns of interaction, sustaining functions, information, or even memes,[1] had long been considered to fall under the scope of what evolutionary biology, or eventually cultural evolution, should intuitively be able to explain, yet struggled to. ITSNTS recently appeared as a powerful way to fill in this important theoretical gap.

When formulating their hypothesis, Doolittle and Inkpen were very careful to highlight both its unique scientific and philosophical

scopes. The key point of ITSNTS is that some processes could now be understood as bona fide units of selection, which are different from recognized higher level entities that emerged by selection through evolutionary transitions,[3] for example, from smaller units giving rise to a higher level biological organization, as genomes would emerge from genes, or as eukaryotes would have emerged from an endosymbiosis between two prokaryotic partners.[4] Such cases are typically accounted for in the context of multi-level selection.[5,6] According to Doolittle and Inkpen, biogeochemical cycles, metabolic functions or other functions realized by communities of individuals, information, and even memes, *qua* processes, can be described as patterns of interactions between components (typically, in their biological examples, such components were entities, such as genes, organisms, or species.[1]) Processes (see glossary) are consequently defined as the dynamic results of interactions between components of a system, and can accordingly be described using networks to represent the patterns of interactions that sustain the targeted process. Such networks might, in many cases, only be proxies for the phenomenon we mean to explain, for example when co-occurrence networks are used to infer molecular or organismal interactions that perform the function of a system.[7] Whereas the Darwinian theory of ENS could explain how replicated components could be selected,[8] until ITSNTS was proposed, a major theory explaining how interactions between components could, eventually, also be the result of natural selection was lacking.[9] To embrace ITSNTS wording, Darwinian ENS successfully explains the evolution of singers (basic individual entities), yet it could not easily explain the sideways evolution of another Domain of selective units: the songs (processes resulting from interactions between potentially heterogeneous individual entities).

In particular, ITSNTS is distinct from ENS, because songs are not made of singers (in the same way that genomes are made of genes), but rather songs are performed by singers.[1] At first sight, this distinction stressing the difference in nature between a song and its components seems to leave little opportunities for songs to expose traits upon which natural selection could act. This traditional reading suggests that only singers may be the real outcomes of evolution by natural selection, whereas songs might be at best a fortuitous evolutionary by-product of selection operating at a narrower level.

To this perspective, Doolittle and Inkpen oppose a process-based approach to evolution, explicitly grounded in process-based metaphysics.[1,10] In this line of thought, songs are processes and can feature the necessary coherence to be conceived as persisting individuals, that is, as defined by ITSNTS, fitness bearers emerging from components in evolutionary dynamics. Doolittle and Inkpen write: "Conversely, process ontologists see processes as primary, [material] things as their manifestations. For ITSNTS, this would mean, in effect, that taxa and the communities they form are adaptations of the processes they implement, not the other way around. And of course we, as multicellular individuals, are processes: few of our cells are "the same" as those with which we were born, and our identity is sustained throughout our lives by the continuity of developmental and regenerative processes, not the atoms or cells in our bodies at birth."[1] Doolittle and Inkpen insist that some songs can be selected for themselves (i.e.,

songs are selectable patterns of interaction).[1] While, they sketched the importance of topological descriptions of interaction structures susceptible to get selected, however, they did not elaborate upon the practical, empirical outcome of their theory for evolutionary studies.

Here, we further develop ITSNTS to provide a method for possible empirical tests for the evolution of patterns of interaction, aiming to distinguish processes that can evolve by natural selection from those that cannot. We agree with Doolittle and Inkpen that any given song (e.g., a geochemical cycle such as nitrogen fixation) is not made of singers but performed by singers. Yet, we feel that it could be even more productive to stress that there is in fact no fundamental ontological divide between singers and songs, in other words that evolution by natural selection is about patterns of interactions, all the way down. Our model, called ITSATS (for It Is The Song *And* The Singers) is more closely related to process-based metaphysics than ITSNTS. Indeed, while ITSNTS holds that both processes (songs) and entities (singers) are mandatory to understand evolution, ITSATS stresses that the whole biological hierarchy can be fruitfully described using patterns of interaction, aligning with contemporary advocates of the relevance of process-based metaphysics for biology. Typically, Nicholson and Dupré wrote: "we propose that the living world is a hierarchy of processes, stabilized and actively maintained at different timescales. We can think of this hierarchy in broadly mereological terms: molecules, cells, organs, organisms, populations, and so on. Although the members of this hierarchy are usually thought of as things, we contend that they are more appropriately understood as processes."[10]

Yet, an important difference remains between ITSATS and traditional process-based metaphysics, which maintains that entities and processes are opposed notions, sustaining mutually exclusive approaches to the description of science's subject matter. ITSNTS maintains the distinction, but rejects the idea that the corresponding views of evolution are mutually exclusive. In contrast, ITSATS claims that embracing a common ontology, according to which *all* evolving entities can be described as patterns of interaction[9] provides a scientifically operational framework, inspired by work in systems biology as well as novel approaches in evolutionary biology.[11-13] More explicitly than ITSNTS, ITSATS thus emphasizes the hierarchical structure of systems, pointing out that at any level objects (singers) interacting to produce a song can be modeled as songs themselves, performed by other interacting objects (singers) at the next level down. Singers and songs become labels to be applied contextually. For example, a singer, such as a unicellular organism part of population dynamics or symbiotic interactions, can become a song in an explanation focused on processes resulting from more inclusive interaction networks.

Also, ITSATS emphasizes how interacting sets of objects (songs) may evolve in various ways, both neutrally and by natural selection operating through persistence rather than reproduction. This allows testing (and possibly rejecting) the ITSNTS explanation of some processes, for instance, when patterns of interaction are not repeated, since then a song is not re-produced. Moreover, interaction patterns can be mathematically characterized, their topological variants can be described in terms of centrality and path analyses,[14-16] which provides a mathematical framework to study the fitness of a song,

BOX 1 Fitness and patterns of interactions in the context of ITSATS

ENS occurs in population where there is variation, differential fitness and heredity. Because ITSATS is meant to augment the scope of evolutionary biology, it must show, among other things, that some patterns of interactions can indeed be fitness bearers, increasing their frequency in a population as a result of selection. We suggest that the fitness of songs is intimately related to the notions of resilience and robustness, which can be illustrated as follows.

First, some biological phenomenon can be represented as a network of interaction. There are countless methods available to achieve this, as the tools of network sciences inform evolutionary biology in a many ways. In the case of microbial communities, for example, co-occurrence networks can be used to infer interaction networks (see Box 2 and the work of Faust and Raes.[50])

Second, once an interaction network has been generated, it is possible to identify patterns of interaction (subgraphs) and measure their fitness, just like this can be done with traditional units of selection. An important distinction, however, is that traditional ENS reduces fitness to the reproductive output of units of selection, while ITSATS takes a more inclusive approach, in which fitness refers to the persistence of a system. Persistence of a system can be realized in two ways: when patterns of interactions are robust and when they are resilient. In Figure 1, we illustrate both phenomena in a multilayer network. Such a multilayer network describes topologies of interaction networks at different time points, each layer being constructed using the method presented in Box 2. Figure 1 shows two different occurrences of the same pattern of interaction (a triangle). One of these (red) is used to illustrate robustness, as the pattern remains the same across time; the other (blue) illustrates resilience, as the pattern is perturbed and then re-produced. Selection on persistent or resilient systems can thus increase the fitness of a pattern of interaction.

Accordingly, the fitness of "traditional organisms" has already been studied through topological considerations on the (functional) interactions of organismal components. This methodological treatment can be found in the literature on ageing, because, precisely, ageing is defined as a decrease in fitness. For instance, Gavrilov and Gavrilova [70] proposed that organismal fitness reduces through a process of redundancy exhaustion, ultimately leading to a system with components connected in series, so that any cumulation of new defects in the components can only lead to organismal death, leveling off mortality rate of the organisms in the meantime. Thus, organismal fitness would benefit from redundant singers (playing the same part in the song) that reduce the chances of having defects in the connected series. Furthermore, Kriete[71] proposed that negative feedbacks between organismal components can introduce some robustness in networks, before non-heritable damage results in catastrophic system failure (death). In other words, specific patterns of interactions can be critical to stabilize organismal fitness. Likewise, Kiss[72] proposed to analyze fitness using the notion of network evolvability, that is, the capacity of the system to change its own patterns of interactions. Beyond robustness, organismal fitness requires some flexibility in the interactions between its singers. Literature on gut microbiome and ageing provides additional examples in which interactions between host-associated microbes are correlated with host fitness (healthy ageing)[73]. Such studies consider the fitness of "traditional" organisms, as resulting from broader, dynamic, and eventually tipping structured interaction networks involving microbes.

that is, the appearance and increased re-production of new interaction patterns over time/space.

Below, we develop the scientific pay-offs of ITSATS modeling.

Both singers and songs can be described as patterns of interaction

ITSNTS stressed that songs are distinct units of selection from the singers performing them. Doolittle and Inkpen wrote that: "whatever biological processes are made of, it is not the same kind of stuff as the things (genes, cells, organisms, species) that implement them."[1] Thus, ITSNTS seems to oppose two kinds of entities: patterns of interactions (the songs) and other entities (the singers), while acknowledging that both kinds can evolve sideways. The relationships between singers and songs are however complicated. On the one hand, Doolittle and Inkpen underscore a strong relationship of dependence between songs and singers. They write that "because there is a song, there are singers,"[1]

and that "because there are singers, there is a song."[1] On the other hand, ITSNTS is rooted in the fundamental possibility of a decoupling between the evolution of singers and songs: the singers could change and the song persists, since songs can be real units of selection. Because these aspects of the ITSNTS hypothesis are complex, they can easily become a source of confusion, in particular because singers (being more traditional objects of studies in evolutionary biology than songs) may still seem to be the more fundamental entities through which to understand evolution, and songs might continue to be overlooked (in evolutionary biology).

We propose therefore, and in contrast with ITSNTS, that it would be fruitful to stress the lack of fundamental ontological distinction between songs and singers. Singers, like songs, can be real units of selection, in the very sense that even the most basic material entities, familiar to biologists, such as genes or organisms can also be seen as emerging from patterns of interactions.

Let's consider a seemingly very basic component of life, an entity to which it seems intuitive to assign the role of singer: a gene. From

a biochemical perspective, a gene can also be seen (and modeled) as the result of chemical, physical and spatial interactions between DNA bases. In that case, according to ITSATS, because a gene re-produces by re-assembly of a collective of DNA bases in interaction, a gene function associated with this pattern of interaction can be selected for. In other words, a gene can also be seen as a song performed by a pattern of interactions between singers (its DNA bases). This example is a direct extension of ITSNTS at the molecular level of biological organization: because there is a given gene function that can be selected for, there will be a specific re-assembly of DNA bases, and because there are specific interactions between DNA bases, there will be a given gene function.

Empirically, this takes notable importance when genes and their functions are generated by the recruitment of genetic material coming from different gene lineages. For example, more than three hundred composite genes were identified in haloarcheal lineages, 126 of which were formed with genetic material coming from bacterial lineages.[17] This process, importantly, is distinct from lateral gene transfer, where complete genes are being transferred from one lineage to the next. In contrast, composite genes arise when smaller DNA segments associate to form new genes, and are said to be phylogenetically chimeric when these genes stem from genetic material issued from distinct host lineages.

Composite genes are, from an empirical standpoint, extremely telling: they highlight the complexity of evolutionary dynamics that give rise to evolutionary novelties such as the adaptation of haloarchea to high salinity environments. From a conceptual standpoint, they are just as important: not only can genes be described as patterns of interactions between bases, but their components are potentially obtained from distinct lineages, such that genes can lean towards having "too many parents."[1] While tracking even smaller bits of DNA, or even lineages of nucleotides, might become tempting in order to compensate for the increasingly recognized messiness of gene lineages,[17–19] this would only transpose the difficulty one level down (genes are songs composed of smaller singers, which are songs composed of yet smaller singers, etc.). In the end, when DNA sequences such as genes are tracked, for example, by phylogeneticists, it is not as material monoliths that they attract attention and yield results, but rather as re-produced selectable sets, that is, patterns of interaction or songs. Hence, gene lineages feature the same kind of stability found in other patterns of interactions that ITSATS identifies as *bona fide* units of reproduction.

Modeling gene evolution with dynamic interaction patterns

There are several pay-offs to consider the function of a gene as a pattern of interaction. First, this view seems consistent with the Dawkinsian model of evolution,[20] a popular description of evolution by natural selection. The Dawkinsian model stresses the success of replicators, that is, lineages of immortal genes, bearing a replicated information. In this model, genes grow into families of selfish genes, through rounds of rather faithful replications. Although the material of the ancestral gene

itself is rapidly replaced by new bases, reassembled according to the initial template, the information (or, according to ITSATS, the song) carried out by that gene persists. Accordingly, it is tempting to summarize the evolution of a replicator and the evolution of its information using an evolutionary tree, which represents how substitutions accumulate within a gene family as new copies of the selfish genes are produced by replication. Such an evolutionary tree is more conventionally called a gene phylogeny. Such a standard tree-based formalism offers a natural way to display and to analyze the evolution of a gene function and is furthermore consistent with the practices of phylogeneticists as well as with the Darwinian and the Dawkinsian views of evolution by natural selection.

However, molecular biologists and phylogeneticists are well aware that such gene trees can only partially capture gene functional evolution.[21] Namely, a gene phylogeny represents the succession of substitutions in the primary sequence of homologous genes over time, which uncontroversially constitutes an essential part of the history of the genes. But from that tree pattern it is impossible to fully predict the function of each gene variant that evolved. The reasons for these are well established. During evolution, some base substitutions can be neutral with respect to the gene function: the standard notion of synonymous substitution illustrates that an "A" can be replaced a "T" in the primary sequence of the gene without affecting the gene function, or more exactly that of its encoded protein(s). By contrast, other mutations can alter the gene function, for example, a non synonymous substitution at a base encoding the active site of a protein might change the function of that gene and thus its song. Yet, this functional knowledge depends on interactions between bases forming a gene, and also on interactions of that gene and its encoded products with other molecular components.[12,22] This means that describing the evolution of functions, or of information, remains a non-trivial issue, which cannot systematically be solved by a simple mapping of changes in a tree-pattern of descent with modification.[9,21]

Interestingly, conceiving of a gene as a pattern of interactions rather than as a singer allows an expansion of the ways to describe the evolution of songs beyond the canonical phylogenetic trees/replicators approaches. For example, when DNA bases enter into novel interaction patterns as a result of a gene fusion, a new song (or function) can evolve and be selected for (gene fusion does not necessarily lead to a new function, but it can).[23,24] This is the case for so-called fused genes with emergent properties, such as the AtGRXS16 S-gene family in *Arabidopsis thaliana*, formed by the fusion of components from the GIY–YIG and GRXS domains, encoding amino-acids between which an intramolecular disulfide bond can be formed, an interaction with functional consequences that would not exist in absence of that gene fusion.[18] Because the origination of songs by fusion involved (at least) two distinct sources (e.g., in the case of AtGRXS16, DNA from two distinct gene families), the evolution of that kind of information cannot by definition be described by a tree. Indeed, trees have single roots, where fused genes phylogenies would require multiple roots.[25] This simple example, coupled with the chimeric genes example provided earlier, reiterates the central tenet of ITSATS, which is that the evolution of a song (e.g., a function sustained by interacting DNA bases) is a more

general issue than the description of the changes occurring on singers (e.g., the substitutions of DNA bases) by a branching pattern. Evolution of novel songs has also to do with the variation of the outcomes of interactions between components, even for entities traditionally conceived as singers.

Moreover, a gene or a gene product interacts with other genes and other genes products. These interactions are well known in systems biology,[12] and described for example in the form of gene regulatory networks,[22,26,27] or protein-protein interaction networks.[28,29] These networks associate genes whose interactions result in a particular re-producible phenotype,[30] and such interaction networks are therefore also songs. More precisely, since the ITSATS model already considered genes – and their products, for example, a folded protein, resulting from a recurrent pattern of interactions between amino-acids under physical, chemical and environmental constraints – as songs, ITSATS sees a gene regulatory network as a larger song made of smaller songs, that is, as an opera performed by specific interactions between songs. Because there is a gene regulatory network (opera), there are genes (smaller songs) that occasionally change interactions, for instance when the promoter of a gene is subjected to a mutation that affects how that gene (or its products) interact with other genes (or other gene products) in interaction networks.[27] We are not eliminating singers here, but we do highlight that every singer is a song with respect to the next level down, while still playing the role of a singer with respect to the next level up, and both roles are critical to the hierarchical view. The status of singer and song is defined by the context being considered. As such, ITSATS appears as a fractal and unifying model relying on a single, shared ontology: patterns of interactions (songs *and* singers).

ITSATS provides a single, empirically testable, ontology for evolution

We have argued that evolution by natural selection, understood as the evolution of songs, that is, patterns of interaction with selectable properties, provides a universal model across all biology – from molecules[12] to ecosystems[31–34] – and points towards a shared network-based ontology to analyze the history of life. Bringing forward this network-based ontology for evolution by natural selection is fruitful for many reasons. First, networks are powerful analytical models. Networks allow for topological analyses of the evolution of songs, for all levels of biological organization. Take for instance the case of a metabolic cycle: a series of reactions (e.g., transformation of A into B, transformation of B into C, etc.) performed by enzymes (e.g., 1, 2, 3, etc.) hosted by microbial cells (Figure 2). This cycle is the direct outcome of interactions between enzymes (biotic components) and substrates. As the classic ITSNTS hypothesis described, this metabolic cycle does not require that all singers come from a monogenic population, for example, the enzymes could be carried by different microbial hosts, as illustrated in cases of metabolic hand-offs.[35] This metabolic cycle is a song, if it is re-produced over time, even though different singers (different enzymes, or enzymes from different hosts for a given step) could

be involved in the different occurrences of the metabolic cycle. This metabolic cycle would even be possibly positively selected for if its frequency increased over time. Moreover, as ITSATS stresses, the topology of that metabolic cycle can be characterized by centrality and path analyses.[14–16]

This latter point brings out the second advantage of the network-based ontology: the evolution of songs by ITSNTS is empirically testable. ITSNTS was originally motivated by the striking observation that some ecosystems show recurrent dynamics that are difficult to explain, such as the debated re-production of microbial communities observed in coastal marine plankton time-series by ecologists,[36] or the reproduction of microbiomes observed in human guts by clinicians and metagenomicians.[37] ITSNTS addressed this challenge by providing a framework that explained such re-occurring natural phenomena, by arguing that the re-production of an entity opens the possibility to consider this entity as a *bona fide* unit of selection.

Differential fitness, along with heredity and variation, is a minimal condition for ENS.[8,38] Traditionally, fitness has been reduced to the reproductive output of units of selection. Populations, in this perspective, evolve by means of natural selection when some of their components reproduce more than others, thereby increasing the ratio of their type (gene family, species, phenotype, etc.). Building upon the work of various authors that have shown this approach to fitness to be highly insufficient to account for biological evolution, even in the context of Darwinian evolution.[1,39–44] ITSNTS argues that patterns of interaction can be fitness bearing entities (songs) even if they do not have traditional reproductive output. In this sense, Doolittle and Inkpen use Bouchard's notion of fitness as persistence: fitness refers to the *persistence of lineages* (or more generally songs) across time. This persistence can be achieved by the sheer stability of a song, or by it being re-produced across time, by different singers. In cases of re-production, that is, recruitment of singers performing a song, there need not be material continuity between the singers of a song. But as Doolittle and Inkpen put it: "This is true but not fatal to ENS (for either memes or community metabolisms) as long as there is some causal connection between populations over time, a relationship between implementations of a process [song] now and in the future such that the former can be seen as necessary for the latter."[1]

In other words, Doolittle and Inkpen banked on the fact that songs can persist through re-production,[41] that is, the recurrence of a pattern of interaction. If this is to be translated empirically, however, more needs to be said on how we can measure song fitness. In Box 1, we show how the notion of fitness applies to patterns of interaction using the notions of network robustness and network resilience within multilayer networks, constructed from longitudinal data. In short, song lineages can persist in two ways: either a given song resists change, it is then said to be robust; or it fails to resist perturbations, but its topology tends to be re-produced across time. Patterns of interactions thus re-produced can be said to be resilient. Importantly, this means that reproduction is a specific case of re-production, which is itself a specific way for persistence of systems to be realized. To take a fitness-as-persistence approach to ENS thusly helps us expand the scope of evolutionary biology in a unified way, centered on patterns
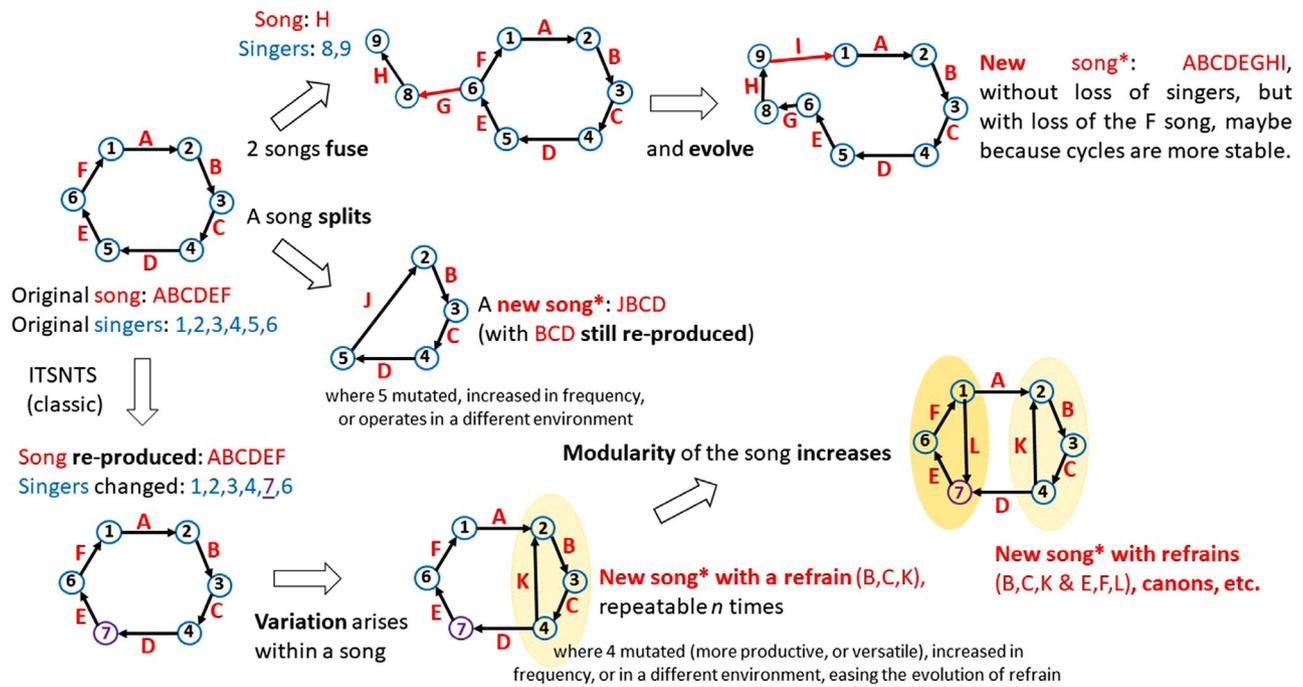
FIGURE 1 Possible schematics for the evolution of the "A,B,C,D,E,F" song. Singers are in blue circles, and their interactions are represented by directed edges. The song corresponds to the succession of red steps along the interaction pattern. For example, component 1 performs the transition from step "F" to step "A," while component 2 performs the transition from step "A" to step "B"; the interactions of components 6 & 1 & 2 therefore plays a part of the complete song: "A,B,C,D,E,F." As the patterns of interaction change, the songs evolve or persist. The evolution of new patterns offers an additional description of evolution, supporting network studies of evolution, according to ITSATS. *refers to case where, if the new song increases in frequency with respect to the "ABCDEF" song, in some environments, then the new song can be positively selected for. Anytime, new song get selected because a singer changed the song, niche construction can occur

of interaction. Indeed, the fitness of a song can be diagnosed by quantifying the frequency at which a given pattern of interaction appears in a population (Figure 1), for example by analyzing –omics data produced in time series. In Box 2, we detail how evolutionary inquiries framed by ITSATS can be realized and Box 3 shows how this can inform contemporary debates in evolutionary biology. These discussions help understand the various ways in which novelty can be introduced into patterns of interactions or, more generally, how songs can change by splitting, or fusing, or by experiencing variants of their patterns of interactions, and more fundamentally songs can change when the interactions between singers change (Figure 2).

Modeling the evolution of ecosystems using dynamic interaction patterns

This opens up many empirical possibilities to test insights that go beyond the work of Doolittle and collaborators, and to frame them in the unifying vocabulary of ITSATS. For example, Lenton and collaborators[45–48] have in the past decades suggested that ecosystems can be units of selection, that is, entities featuring traits that could be selected for and thusly form populations that evolve by natural selection.

This claim has been put to the test recently by Blouin and collaborators[7,49] who designed experiments of artificial selection to act on microbial communities standing in as ecosystems. Specifically, Blouin et al. recently highlighted that CO_2 emissions of microbial ecosystems could be artificially selected. Interestingly, this ecosystemic trait was correlated with patterns of interactions between the constitutive microbes. "A first objective of this study was to bring an experimental proof of principle that community structure, especially the structure of interaction networks of communities, are significantly affected during the artificial selection procedure. A second objective was to document how far we can go in changing ecosystem phenotype by artificial selection." [7] In order to achieve this, their experiments featured six independent lines of thirty microbial communities each (a control set of microbial communities similarly structured was also used to contrast artificial selection with random selection). The communities were left alone for twenty-four hours and their CO_2 emissions were then measured. In each independent line, the three communities that produced the least CO_2 were selected on this basis, pooled together to produce a source from which thirty new communities were created. More precisely, each new community is created from a sample of $50\,\mu$L of the source pool to which $750\,\mu$L of sterile liquid medium was added; and the communities were cultivated in 96-deep-well microplates.

After twenty such selective events (twenty-one generations), the biodiversity of the communities was characterized on the basis of the
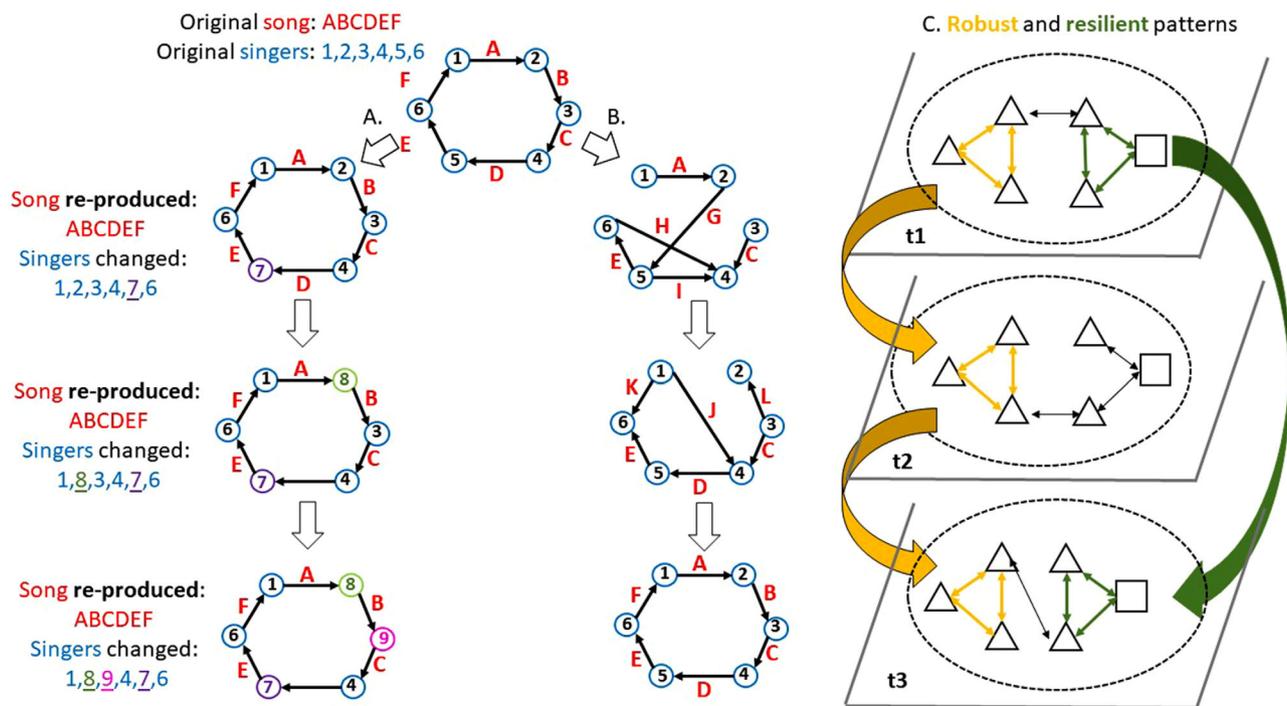
FIGURE 2 Further description of songs' properties. Same colour code as Figure 2. The frequency of recurrence of the song provides a basis to determine (A) whether a song is robust to the change of singers (singers switch a lot, the song stays) or, inversely, (B) whether a song is fragile (Singers are there, rarely the song). (C) Formal modeling using a multilayered network, where each layer (gray area) corresponds to the state of a system (delimited by a parallelogram) at a given time *t*. Nodes are (biotic or abiotic) components of the system, and the two-ways arrows that link them represent interactions (e.g., inferred from co-occurrence networks). This specific multilayered network was built to illustrate the resilience and robustness of patterns of interactions. In orange, a triangular pattern of interaction remains the same across time; it is robust. In green is another occurrence of a triangular pattern of interactions that features resilience: it disappears in the second layer, but is re-produced in the third. In this light, maximizing robustness and re-producibility are two ways for songs (in orange and in green), and for a system, to maximize their fitness

T-RFLP-defined genetic units in presence. This characterization was then used to infer interaction networks (as in Faust and Raes,[50]) correlating lower CO_2 emissions with structural features of the communities. Notably, low CO_2 emission was thusly correlated to low interaction richness (i.e., the total amount of pairwise interactions), low connectance (i.e., the ratio for realized to possible interactions) and low average degree (i.e., the average amount of interaction of nodes present in the network).

This insightful work can be interpreted constructively in the terms of ITSATS: ecosystems are patterns of interactions, which can be acted upon by selection. As in ITSNTS, this statement implies that ecosystems (songs) can evolve by means of natural selection, but it *does not* deny the possibility of the singers, in this case the microbes, to evolve sideways. On the contrary, both evolutionary dynamics appear to be complementary, in the sense that both must be taken into account if biodiversity is to be properly assessed and its underlying dynamics understood.

Basically, ITSATS means that a theoretical framework is now available to mobilize the inferential power of network sciences in ways that could increase our knowledge of evolutionary dynamics. In MLNs, analyses of community structure, and in other uses of network-based tools in evolutionary biology, network comparison holds the keys to a more inclusive understanding of evolution. It becomes testable for instance that songs follow general topological principles over evolutionary time (e.g., an increase in modularity,[14]) if the topology of the past interac-

tion networks differs from the topology of subsequent interaction networks, as the result of introduction of particular motifs over time.[12,51] It is also possible and testable that evolving processes fell prey to the evolution of other competing processes. If new interactions become favored by selection (e.g., step A being no longer followed by step B, but being more frequently followed by step C, so the song lyrics evolves from "AB" to "AC"), then a song can be re-oriented in a novel evolutionary direction.

ITSATS can model the evolution of interconnected processes

A third pay-off of stressing the common network-based ontology to describe evolution by natural selection is that it makes this modeling especially inclusive. As Doolittle and Inkpen noted when they focused on the nature of the singers, songs can be performed by biotic and abiotic components in interactions.[1] This diversity allowed ITSNTS to model the evolution of ecosystems, whereas the inclusion of abiotic material, which by definition does not replicate, as a part of an evolving system usually makes the notion of evolution of the ecosystem as a whole irrelevant for traditional evolutionary theories.[41] Focusing on interactions between *and* within singers, ITSATS proves that ITSNTS can be understood as being even more inclusive.

BOX 2 Investigating the origins of symbioses with the ITSATS framework

How to analyze songs practically? What type of data do we need? What does it bring compared to a more traditional approach? A thought example illustrates what could currently be achieved to shed light on the origins of symbioses, a distinct research avenue from traditional phylogenetic studies on the origins of species.

Patterns of prokaryotic interactions over time (featuring some microbial songs) can be inferred using state-of-the-art microbial co-occurrence networks construction methods. For example, using 16S environmental data from time series, such as a series of seawater samples, receiving increasing amounts of freshwater from the melting ice in Arctic sea, one can build networks, in which nodes correspond to OTUs and edges correspond to statistically significant, weighted correlations in OTUs distribution after a given duration of fresh and sea waters mixing. Producing such temporal networks requires several steps.

In brief, first, sequence reads (from transcriptomics or metagenomics) must be denoised, filtered to remove chimera, and clustered into sequence variants to generate OTUs (for example using DADA2 v1.1.5.[74]) Second, for samples from the same time point (15 samples being considered as a decent starting point), OTUs abundancy tables can be analyzed using Sparse Inverse Covariance Estimation for Ecological Association and Statistical Inference v0.1.2 (SPIEC-EASI)[75] to compute correlation strength, and/or using Weighted Gene Co-expression Network Analysis (WGCNA), to compute co-occurrence patterns (e.g., Maffei et al. 2017.[76])

The resulting temporal series of microbial co-occurrence networks tracks how and which OTUs interact over time. These networks may feature robust interactions (persistent sets of edges and nodes in networks from consecutive sampling times) or resilient interactions (re-produced sets of edges and nodes, in networks from disjoint sampling times; see Box 1). Network comparisons to identify statistically significant matching subgraphs between pairs of microbial co-occurrence networks can take further advantage of edge weights to test for general or local reinforcement dynamics of interactions. For instance, at t_1, when freshwater first meets sea water, the resulting mix of microbial communities may be described by microbial co-occurrence networks, with a diversity of weakly correlated interactions, whereas when the mixing of waters have been ongoing for a longer time period, if the microbial communities become more structured, their microbial co-occurrence networks will present some stronger correlations, a different distribution of edge weights, and the recurrence or persistence of some sets of edges and nodes past time t_n. Thus, emerging microbial symbioses (or more generally any emerging patterns of interactions) could be identified by finding subgraphs that become robust with time (i.e., are present in the co-occurrence networks past t_n, possibly with increasing edge weights, and representing an increasing proportion, or absolute number, of the reads present in each sample). Such reinforced patterns of microbial interactions, suggestive of microbial symbioses, would be compatible with a form of selection, acting beyond a single OTU level. Analyses designed to impose selective constraints on ecosystems (see main text) could further demonstrate causal connections between a given selection regime and the emergence of reinforced patterns of microbial interactions (songs). By showing that some symbioses can be selected, such analyses would go beyond traditional approaches that tend to analyze OTUs evolution independently, and test ITSNTS, that is, whether different sets of microbes (singers) can take part into similar songs.

First, our approach can reconcile neutral evolution,[52–54] ITSNTS and evolution by natural selection. The evolution of interactions is indeed central for the evolution of songs. According to ITSATS, re-produced interactions can also be seen as songs, possibly within a larger song. But critically, not all interactions are the result of natural selection. Constructive neutralism (also called pre-suppression)[52–55] explains how some interactions can (irremediably) arise in absence of positive selection for them. For example, the complexity of the ribosomal machinery may first evolve neutrally, as a result of pre-suppressed mutations introducing dependences between ribosomal proteins.[56] Likewise, dependency relationships between microbes are expected to evolve by a ratchet mechanism in nature.[55]

Typically, some biological interactions likely arise by chance: by pre-suppression or even by drift (within populations with small effective sizes).[52] This kind of evolution probably typically affects gene regulatory networks. Thus, "the regulator first hypothesis" holds that many components of a gene regulatory network are neutrally recruited by association with the recruitment of a promoter, rather than based on particular selective advantages carried by each component of a gene regulatory network.[57] In this model, neutral connections are not counter-selected, new edges are expected to accumulate neutrally in gene regulatory networks, producing "fatter" patterns of interaction (longer songs) than expected by chance alone. Consequently, gene regulatory networks feature a mix of connections, some of which are essential for their new functions, whereas some other connections (the regulated genes and edges recruited by association with a newly acquired promoter) are a neutral, by-product of the mode of network growth.

Just like Doolittle and Inkpen introduced abiotic components as *bona fide* singers of songs, ITSATS introduces neutrally evolved songs (neutrally evolved interactions) as *bona fide* components of operas. What seemed like a non-starter for evolution by natural selection (constructive neutralism) becomes an acceptable evolutionary process, able to produce stable, re-producible patterns of interaction between

---

BOX 3 ITSATS informs contemporary debates in evolutionary biology – the case of host-mircobes interactions

ITSATS also informs contemporary debates in evolutionary biology.

For example, the evolutionary potential of symbioses generated by interactions between a macrobial host and its associated microbes is currently a hotly debated issue.[77–83] While multispecies symbiotic communites that include a macrobe have been shown to be functionally integrated in ways that influence the components' fitness, theoretical arguments have been proposed to deny most of these symbiotic assemblages the status of unit of selection because of their lack of fitness-based unity[79,84] or their incapacity to reproduce as wholes.[5] Other authors have defended the opposite, banking on the interactor/replicator distinction[81,82] or by stressing the functional continuity of such assemblages across generations.[85] Doolittle and Inkpen, with ITSNTS, highlighted the possibility that some of such assemblages are songs, and hence units of selection. Most empirical models that were developed to assess the evolutionary potential of these consortia, as seen in the work of Roughgarden,[81,86] describe host-microbe interactions on the basis of taxonomic composition, a promising approach in itself, but that leaves aside the study of the interactions among these taxa.

ITSATS provides another line of empirical inquiry to assess the evolutionary potential of host-microbes associations. For example, co-occurrence networks (see Box 2) of constitutive singers (hosts and microbes) could be used to infer if some patterns of interaction qualify as specific candidate holobionts. Importantly, this requires establishing criteria for comparing networks of interactions, such as homomorphy, connectance, average degree of nodes composing the network, and so forth. This type of inquiry can test hypotheses regarding structures of interactions deemed more fit than others, and could highlight topological traits of host-microbe associations that are targeted by selective pressures. It could, for example, stress the centrality/preferential connections of few nodes in the network and hence weaken the case for whole community-level selection, yet identify critical subsets of robust or resilient interactions between hosts and microbes (i.e., smaller holobionts).

This echoes recent work by Suárez,[85] who suggests that host-microbes symbioses are evolutionary relevant, if their boundaries are set by identifying the microbial genes that contribute, through interactions, to the persistence of the symbiosis across time. In his approach, the whole macrobial genome is included in the functionally delineated symbiosis, while only essential genes of microbes are to be considered. This perspective could be strengthened by demonstrating which macrobial and microbial genes are central in the patterns of interaction that characterize host-microbe symbioses, thereby acknowledging their privileged role for explaining the resilience and robustness (i.e., fitness) of the interactions at work. ITSATS offers a methodology to test such hypotheses. Alternative network analysis methods (e.g., Boolean network analysis,[87]) might also be used to analyze the dynamics of interactions.

Thus, ITSATS allows to assess the evolutionary potential of the *interactions* that sustain symbioses rather than the evolutionary potential of the taxonomic collections that compose them. By providing this additional outlook, ITSATS opens up the possibility to identify shared topological structures, present in different putative holobionts in the form of conserved networks of interactions composed of singers.

---

components (songs), embedded in a broader picture of evolution (operas). For example, the complexity of the ribosomal machinery may first have evolved neutrally, but because ribosomes contribute to perform a step of a larger re-produced cycle, the cell cycle, since there is a cell cycle, a complex ribosomal machinery is also re-produced. A neutrally constructed song can be co-opted as part of a selected opera.

Second, modeling the evolution of operas, as allowed by ITSATS, is desirable to understand the evolution of processes in a realistic way, since processes are typically interconnected. Not only can larger songs be made of smaller songs, but songs can also intersect as shown in countless examples in ecology and in systems biology.[58–60] Niche construction[61] is a neat example for such interconnectedness with evolutionary significance. When the outcome of a metabolic network (song 1) is a component, used in an interaction to perform a step of a geochemical cycle (song 2), then a song (the metabolic pathway) contributes to another larger song (the geochemical cycle), which in turn contributes, by niche construction, to the success or failure of the singers (enzymes and host taxa) that performed the metabolic song. Thus, a metabolism like denitrification, producing N_2 thanks to

a final reaction involving the gene *nosZ* will ultimately enhance the fitness of the *nosZ* gene if the Nitrogen cycle, of which N_2 is a component, is fully realized. However, if microbes use the Annamox variant path, from NO_2^- to N_2, within the Nitrogen cycle, this alternative path (i.e., song) would favor the *hzo* gene over the *nirS/K*, *norB* and *nosZ* genes[62] (Figure 3). In that sense, ITSATS makes it possible to model how interconnected pathways (which may have appeared at different geological ages) may compete in different environments. Importantly, it adds perspectives based on -omics approaches (very common in biology) and evolutionary approaches (very fundamental in biology) to the study of processes such as elemental cycles, as well as to the ever-more present eco-evolutionary approaches to the study of various phenomena.[63–69]

The ontology of evolution as patterns of interactions promoted by ITSATS thus brings evolutionary biologists way beyond the classic ENS model, since, according to Darwin: "natural selection cannot possibly produce any modification in any one species exclusively for the good of another species,"[2] and encourages them to analyze the interconnectedness of pathways.
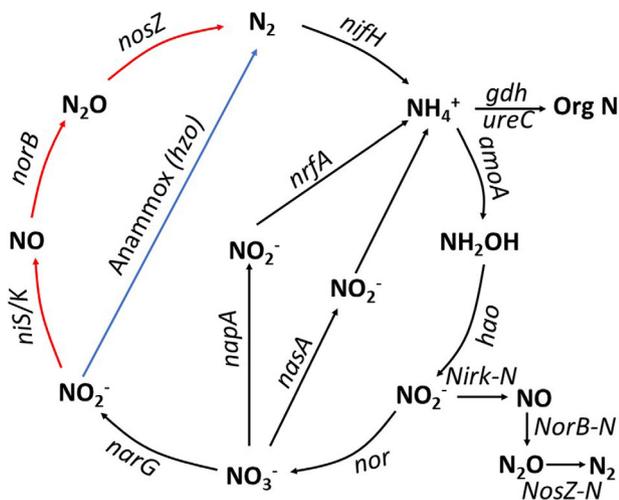
FIGURE 3 Simplified representation of gene-based interactions and phylogenetic diversity in the Nitrogen cycle. The cycle is inspired from Carter et al. 2012. While this representation does not take into account thermodynamic drivers of the cycle, it allows for -omics based comparative studies of the gene sets involved in N cycling over space and time. Components of the Nitrogen cycle are indicated in bold. Genes performing the different steps of the cycles are indicated in italics. Denitrification (a song) is underscored in red, Annamox (another song) is underscored in blue. These two songs may compete with one another. The realization of Denitrification rather than the Annamox pathway may result in niche construction, when N_2 production favors the completion of the Nitrogen cycle, which in turns enhances the fitness of organisms carrying, for example the *nosZ* gene

CONCLUSIONS AND OUTLOOK

None of the above is at odds with the ITSNTS hypothesis. However, ITSATS makes it more salient that inclusive evolutionary studies are possible when evolution is fundamentally analyzed, and fruitfully so, as stabilized (or unstable) dynamics of patterns of interaction. By stressing that singers are not fundamentally different from songs, when singers are considered at a finer analytical grain unraveling their fundamental processual nature, then singers appear to be themselves decomposable into smaller musical parts. Hence, ITSATS posits that larger songs, larger delineated patterns of components in interactions, are composed by the interplay of smaller songs, because components of the larger songs (a.k.a. singers in ITSNTS) are themselves the result of patterns of interactions between smaller components. This precision matters: ITSATS thus underscores how diverse and complex the phenomena that could evolve by natural selection are. For example, it allows evolutionary biologists to embrace, under the legitimate scope of evolutionary explanations by natural selection, the evolution of "operas" as re-produced interaction patterns involving components, which are themselves processes, and the evolution of partly neutrally emerging collectives. Moreover, ITSATS provides a description of this complex evolution currently amenable to scientific studies based on network analyses. Therefore, we invite evolutionary biologists to construct such networks in order to analyze the interactions in the systems on which they work, and, 160 years after the Origin of Species,

Glossary

It's the song, not the singer (ITSNTS): a theory developed by Doolittle and collaborators that is meant to augment the scope of evolution by means of natural selection, claiming that it applies to patterns of interactions (processes) *as well as* to traditional objects of evolutionary biology, such as genes or organisms.

It's the song *and* the singer (ITSATS): the theory we defend in this paper; it is an extension of ITSNTS. We argue that patterns of selection can indeed be selected for, and that this allows one to model the evolution not only of songs, but also of singers. ITSATS defends a unified view of evolution, downplaying the importance of the distinction between songs and singers, and proposes that network methods offer a generic way to assess, empirically, the evolution of patterns of interaction.

Songs: as defined in this paper, songs are patterns of interaction that are also fitness bearers. Such patterns of interactions form lineages whose evolution is sustained by natural selection. Genes, geochemical cycles, multispecies symbiotic assemblages, and so forth. can be described as patterns of interaction, and their fitness can be measured based on their robustness or resilience.

Singers: as defined in this paper, singers are the components of songs. A gene, for example, can be a singer with respect to a larger song such as an organism or a multispecies assemblage. The same gene, in an alternate situation, will itself be considered a song, as it is also a fitness-bearing process resulting from the interaction of its components (domains, nucleotides, etc.), that is, a pattern of interaction that features selectable resilience and/or robustness.

Process: dynamic phenomena sustained by interactions of components, that is, a pattern of interaction. Typically, functions, mechanisms as well as complex entities can be conceived as processes. Processes can be described based on the interactions that sustain them and accordingly modelled using networks. The philosophical notion of process goes back to Heraclitus (535-475 B.C.) and is rooted in a rich philosophical tradition. Recently, its usefulness for biology has been underscored to draw the attention of researchers towards neglected objects of inquiries (namely, in the case of evolutionary biology, patterns of interactions that sustain phenomena of interest).

Fitness: as defined in this paper, the fitness of a biological entity is its capacity to sustain or increase the ratio of similar entities within an evolving population. When applied to patterns of interactions, fitness is translated by the network's robustness (its capacity to resist change) and resilience (its capacity to be re-produced).

purposely tackle the broader and deeper issue of the evolution of processes.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

É.B & FP wrote the MS and produced Figure 1. EB produced the other figures.

ADDITIONAL INFORMATION

No datasets were generated or analyzed during the current study.

No computer codes or algorithms were used to generate results.

ORCID

François Papale https://orcid.org/0000-0003-0641-4645

REFERENCES

1. Doolittle, W. F., & Inkpen, S. A. (2018). Processes and patterns of interaction as units of selection: An introduction to ITSNTS thinking. *Proc. Natl. Acad. Sci. USA*, *115*, 4006–4014.
2. Darwin, C. A. (1859). *On the origin of species by means of natural selection*. London: John Murray.
3. Szathmáry, E. (2015). Toward major evolutionary transitions theory 2.0. *Proc. Natl. Acad. Sci. USA*, *112*, 10104–10111.
4. McInerney, J. O., O'Connell, M. J., & Pisani, D. (2014). The hybrid nature of the Eukaryota and a consilient view of life on Earth. *Nat. Rev. Microbiol.*, *12*, 449–455.
5. Godfrey-Smith, P. (2015). Reproduction, symbiosis, and the eukaryotic cell. *Proc. Natl. Acad. Sci. USA*, *112*, 10120–10125.
6. Michod, R. E. (2006). On the transfer of fitness from the cell to the multicellular organism. *Biol. Philos.*, *20*, 967–987.
7. Blouin, M., Karimi, B., Mathieu, J., Lerch, T. Z., & Grover, J. (2015). Levels and limits in artificial selection of communities. *Ecol. Lett.*, *18*, 1040–1048.
8. Lewontin, R. C. (1970). The units of selection. *Annu. Rev. Ecol. Syst.*, *1*, 1–18.
9. Bapteste, E., & Huneman, P. (2018). Towards a dynamic interaction network of life to unify and expand the evolutionary theory. *BMC Biol.*, *16*, https://doi.org/10.1186/s12915-018-0531-6.
10. Nicholson, D. J., & Dupré, J. (2018). *Everything flows: Towards a processual philosophy of biology*. Oxford: Oxford University Press.
11. Oltvai, Z. N., (2002). Systems biology: Life's complexity pyramid science. *298*, 763–764.
12. Alon, U. (2007). *An introduction to systems biology: Design principles of biological circuits*. Boca Raton: Chapman & Hall/CRC.
13. Watson, A. K., Habib, M., & Bapteste, E. (2020). Phylosystemics: Merging phylogenomics, systems biology, and ecology to study evolution. *Trends Microbiol.*, *28*, 176–190.
14. Newman M. E. J. (2018). *Networks*. Oxford: Oxford University Press.
15. Strogatz, S. H. (2001). Exploring complex networks. *Nature*, *410*, 268–276.
16. Corel, E., Lopez, P., Méheust, R., & Bapteste, E. (2016). Network-thinking: Graphs to analyze microbial complexity and evolution. *Trends Microbiol.*, *24*, 224–237.
17. Méheust, R., Watson, A. K., Lapointe, F.-J., Papke, R. T., Lopez, P., & Bapteste, E. (2018). Hundreds of novel composite genes and chimeric genes with bacterial origins contributed to haloarchaeal evolution. *Genome Biol.*, *19*, 1–12.
18. Méheust, R., Zelzion, E., Bhattacharya, D., Lopez, P., & Bapteste, E. (2016). Protein networks identify novel symbiogenetic genes resulting from plastid endosymbiosis. *Proc. Natl. Acad. Sci. USA*, *113*, 3579–3584.
19. Méheust, R., Bhattacharya, D., Pathmanathan, J. S., McInerney, J. O., Lopez, P., & Bapteste, E. (2018). Formation of chimeric genes with essential functions at the origin of eukaryotes. *BMC Biol.*, *16*, 30. https://doi.org/10.1186/s12915-018-0500-0
20. Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
21. Bapteste, E., & Dupré, J. (2013). Towards a processual microbial ontology. *Biol. Philos.*, *28*, 379–404.
22. Britten, R. J., & Davidson, H. (1969). Gene regulation for higher cells: A theory. *Science*, *165*, 349–357.
23. Enright, A. J., & Ouzounis, C. A. (2001). Functional associations of proteins in entire genomes by means of exhaustive detection of gene fusions. *Genome Biol. 2*, research0034.1.
24. Song, N., Joseph, J. M., Davis, G. B., & Durand, D. (2008). Sequence similarity network reveals common ancestry of multidomain proteins. *PLoS Comput. Biol.*, *4*, e1000063.
25. Haggerty, L. S., Jachiet, P.-A., Hanage, W. P., Fitzpatrick, D. A., Lopez, P., O'Connell, M. J., … McInerney, J. O. (2014). A pluralistic account of homology: Adapting the models to the data. *Mol. Biol. Evol.*, *31*, 501–516.
26. Koch, C., Konieczka, J., Delorey, T., Lyons, A., Socha, A., Davis, K., … Roy, S. (2017). Inference and evolutionary analysis of genome-scale regulatory networks in large phylogenies. *Cell Syst.*, *4*, 543–558.e8.
27. Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature*, *430*, 85–88.
28. Gerke, M., Bornberg-Bauer, E., Jiang, X., & Fuellen, G. (2006). Finding common protein interaction patterns across organisms. *Evol. Bioinforma.*, *2*, 117693430600200.
29. Zitnik, M., Sosič, R., Feldman, M. W., Leskovec, J. (2019). Evolution of resilience in protein interactomes across the tree of life. *Proc. Natl. Acad. Sci. USA*, *116*, 4426–4433.
30. Cassar, P. A., Carpenedo, R. L., Samavarchi-Tehrani, P., Olsen, J. B., Park, C. J., Chang, W. Y., … Stanford, W. L. (2015). Integrative genomics positions as a novel ribonucleoprotein within the embryonic stem cell gene regulatory network. *EMBO Rep.*, *16*, 1334–1357.
31. Chaffron, S., Rehrauer, H., Pernthaler, J., & von Mering, C. (2010). A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res.*, *20*, 947–959.
32. Faust, K., Lima-Mendez, G., Lerat, J.-S., Sathirapongsasuti, J. F., Knight, R., Huttenhower, C., … Raes, J. (2015). Cross-biome comparison of microbial association networks. *Front. Microbiol.*, *6*, https://doi.org/10.3389/fmicb.2015.01200.
33. Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., … Raes, J. (2015). Ocean plankton. Determinants of community structure in the global plankton interactome. *Science*, *348*, 1262073.
34. Karimi, B., Maron, P. A., Chemidlin-Prevost Boure, N., Bernard, N., Gilbert, D., & Ranjard, L. (2017). Microbial diversity and ecological networks as indicators of environmental quality. *Environ. Chem. Lett.*, *15*, 265–281.
35. Castelle, C. J., & Banfield, J. F. (2018). Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell*, *172*, 1181–1197.
36. Martin-Platero, A. M., Cleary, B., Kauffman, K., Preheim, S. P., McGillicuddy, D. J., Alm, E. J., & Polz, M. F. (2018). High resolution time series reveals cohesive but short-lived communities in coastal plankton. *Nat. Commun.*, *9*, 266.

37. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., & Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature*, *489*, 220–230.

38. Godfrey-Smith, P. (2009). *Darwinian populations and natural selection*. New York: Oxford University Press.

39. Bouchard, F. (2008). Causal processes, fitness, and the differential persistence of lineages. *Philos. Sci.*, *75*, 560–570.

40. Bouchard, F. (2011). Darwinism without populations: A more inclusive understanding of the "Survival of the Fittest". *Stud. Hist. Philos. Sci. Part C Stud. Hist. Philos. Biol. Biomed. Sci.*, *42*, 106–114.

41. Bouchard, F. (2014). Ecosystem evolution is about variation and persistence, not populations and reproduction. *Biol. Theory*, *9*, 382–391.

42. Charbonneau, M. (2014). Populations without reproduction. *Philos. Sci.*, *81*, 727–740.

43. Bourrat, P. (2014). From survivors to replicators: Evolution by natural selection revisited. *Biol. Philos.*, *29*, 517–538.

44. Bourrat, P. (2015). Levels, time and fitness in evolutionary transitions in individuality. *Philos. Theory Biol.*, *7*, 1–17.

45. Lenton, T. J. (1998). Gaia and natural selection. *Nature*, *394*, 439–447.

46. Williams, H. T. P., & Lenton, T. J. (2008). Environmental regulation in a network of simulated microbial ecosystems. *Proc. Natl. Acad. Sci. USA*, *105*, 10432–10437.

47. Lenton, T. M., Daines, S. J., Dyke, J. G., Nicholson, A. E., Wilkinson, D. M., & Williams, H. T. P. (2018). Selection for gaia across multiple scales. *Trends Ecol. Evol.*, *33*, 633–645.

48. Latour, B., & Lenton, T. M. (2019). Extending the domain of freedom, or why gaia is so hard to understand. *Crit. Inq.*, *45*, 659–680.

49. Raynaud, T., Devers, M., Spor, A., & Blouin, M. (2019). Effect of the reproduction method in an artificial selection experiment at the community level. *Front. Ecol. Evol.*, *7*, 416.

50. Faust, K., & Raes, J. (2012). Microbial interactions: From networks to models. *Nat. Rev. Microbiol.*, *10*, 538–550.

51. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, *298*, 824–827.

52. Brunet, T. D. P., & Doolittle, W. F. (2018). The generality of constructive neutral evolution. *Biol. Philos.*, *33*, 2.

53. Gray, M. W., Lukes, J., Archibald, J. M., Keeling, P. J., & Doolittle, W. F. (2010). Cell biology. Irremediable complexity? *Science*, *330*, 920–921.

54. Stoltzfus, A. (1999). On the possibility of constructive neutral evolution. *J. Mol. Evol.*, *49*, 169–181.

55. Selosse, M. A., Bessis, A., & Pozo, M. J. (2014). Microbial priming of plant and animal immunity: Symbionts as developmental signals. *Trends Microbiol.*, *22*, 607–613.

56. Lukes, J., Archibald, J. M., Keeling, P. J., Doolittle, W. F., & Gray, M. W. (2011). How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life*, *63*, 528–537.

57. Sorrells, T. R., & Johnson, A. D. (2015). Making sense of transcription networks. *Cell*, *161*, 714–723.

58. Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., … Banfield, J. F. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.*, *7*, 13219.

59. Falkowski, P. G. (2015). *Life's engines: How microbes made earth habitable*. Princeton: Princeton University Press.

60. Stüeken, E. E., Kipp, M. A., Koehler, M. C., & Buick, R. (2016). The evolution of Earth's biogeochemical nitrogen cycle. *Earth-Sci. Rev.*, *160*, 220–239.

61. Odling-Smee, J., Laland, K., & Feldman, M. (2003). *Niche construction: The neglected process in evolution*. Princeton: Princeton University Press.

62. Carter, M. Q., Xue, K., Brandl, M. T., Liu, F., Wu, L., Louie, J. W., … Zhou, J. (2012). Functional metagenomics of Escherichia coli O157:H7 interactions with Spinach indigenous microorganisms during biofilm formation. *PLoS One*, *7*, e44186.

63. Wade, M. J. (2003). Community genetics and species interactions. *Ecology*, *84*, 583–585.

64. Yoshida, T., Jones, L. E., Ellner, S. P., Fussmann, G. F., & Hairston, N. G. (2003). Rapid evolution drives ecological dynamics in a predator–prey system. *Nature*, *424*, 303–306.

65. Hendry, A. P. (2017). *Eco-evolutionary dynamics*. Princeton: Princeton University Press.

66. Burak, M. K., Monk, J. D., & Schmitz, O. J. (2018). Eco-evolutionary dynamics: The predator-prey adaptive play and the ecological theater. *Yale J. Biol. Med.*, *91*, 481–489.

67. Le Roux, F., &Blokesch, M. (2018). Eco-evolutionary dynamics linked to horizontal gene transfer in Vibrios. *Annu. Rev. Microbiol.*, *72*, 89–110.

68. Ashby, B., Iritani, R., Best, A., White, A., & Boots, M.(2019). Understanding the role of eco-evolutionary feedbacks in host-parasite coevolution. *J. Theor. Biol.*, *464*, 115–125.

69. Loeuille, N. (2019). Eco-evolutionary dynamics in a disturbed world: Implications for the maintenance of ecological networks. *F1000Research*, *8*, 97.

70. Gavrilov, N., & Gavrilova, S. (2001). The reliability theory of aging and longevity. *J. Theor. Biol.*, *213*, 527–545.

71. Kriete, A. (2013). Robustness and aging—A systems-level perspective. *Biosystems*, *112*, 37–48.

72. Kiss, H. J. M., Mihalik, Á., Nánási, T., Őry, B., Spiró, Z., Sőti, C., & Csermely, P. (2009). Ageing as a price of cooperation and complexity: Self-organization of complex systems causes the gradual deterioration of constituent networks. *Bioessays*, *31*, 651.

73. Tuikhar, N., Keisam, S., Labala, R. K., Imrat Ramakrishnan, P., Arunkumar, M. C., Ahmed, G., … Jeyaram, K. (2019). Comparative analysis of the gut microbiota in centenarians and young adults shows a common signature across genotypically non-related populations. *Mech. Ageing Dev.*, *179*, 23–35.

74. Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P.(2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods*, *13*, 581–583.

75. Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., & Bonneau, R. A., (2015). Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.*, *11*, e1004226.

76. Maffei, V. J., Kim, S., Blanchard, E., Luo, M., Jazwinski, S. M., Taylor, C. M., & Welsh, D. A. (2017). Biological aging and the human gut microbiota. *J. Gerontol. Ser. A*, *72*, 1474–1482.

77. Bordenstein, S. R., & Theis, K. R., (2015). Host biology in light of the microbiome: Ten principles of holobionts and hologenomes. *PLoS Biol.*, *13*, e1002226.

78. Moran, N. A., & Sloan, D. B. (2015). The hologenome concept: Helpful or hollow? *PLoS Biol.*, *13*, e1002311.

79. Douglas, A. E., & Werren, J. H. (2016). Holes in the hologenome: Why host-microbe symbioses are not holobionts. *mBio*, *7*, e02099.

80. Rosenberg, E., & Zilber-Rosenberg, I. (2018). The hologenome concept of evolution after 10 years. *Microbiome*, *6*, 78. https://doi.org/10.1186/s40168-018-0457-9.

81. Roughgarden, J., Gilbert, S. F., Rosenberg, E., Zilber-Rosenberg, I., & Lloyd, E. A. (2018). Holobionts as units of selection and a model of their population dynamics and evolution. *Biol. Theory*, *13*, 44–65.

82. Lloyd, E. A. (2018) Holobionts as units of selection: Holobionts as interactors, reproducers, and manifestors of adaptation. In: S.B. Gissis,E. Lamm,A. Shavit (Eds.), *Landscapes of collectivity in the life sciences* (pp. 351–368). Cambridge: MIT Press.

83. Theis, K. R., Dheilly, N. M., Klassen, J. L., Brucker, R. M., Baines, J. F., Bosch, T. C. G., … Bordenstein, S. R. (2016). Getting the hologenome concept right: An eco-evolutionary framework for hosts and their microbiomes. *mSystems*, *1*, e00028.

84. Bourrat, P., & Griffiths, P. E. (2018). Multispecies individuals. *Hist. Philos. Life Sci.*, *40*, 33.

85. Suárez, J. (2020). The stability of traits conception of the hologenome: An evolutionary account of holobiont individuality. *Hist. Philos. Life Sci.*, *42*, 11.

86. Roughgarden, J. (2019). Holobiont evolution: Model with vertical vs. horizontal microbiome transmission. *BioRxiv*, https://doi.org/10.1101/465310.

87. Huitzil, S., Sandoval-Motta, S., Frank, A., & Aldana, M. (2018). Modeling the role of the microbiome in evolution. *Front. Physiol.*, *9*, 1836.

CrossMark

Towards a Dynamic Interaction Network of Life to unify and expand the evolutionary theory

Eric Bapteste[1,2]* and Philippe Huneman[3]

Abstract

The classic Darwinian theory and the Synthetic evolutionary theory and their linear models, while invaluable to study the origins and evolution of species, are not primarily designed to model the evolution of organisations, typically that of ecosystems, nor that of processes. How could evolutionary theory better explain the evolution of biological complexity and diversity? Inclusive network-based analyses of dynamic systems could retrace interactions between (related or unrelated) components. This theoretical shift from a Tree of Life to a Dynamic Interaction Network of Life, which is supported by diverse molecular, cellular, microbiological, organismal, ecological and evolutionary studies, would further unify evolutionary biology.

Keywords: Evolutionary biology, Interactions, Theoretical biology, Tree of Life, Web of Life

Deciphering diversity through evolution

The living world is nested and multilevel, involves multiple agents and changes at different timescales. Evolutionary biology tries to characterize the dynamics responsible for such complexity to decipher the processes accounting for the past and extant diversity observed in molecules (namely, genes, RNA, proteins), cellular machineries, unicellular and multi-cellular organisms, species, communities and ecosystems. In the 1930s and 1940s, a unified framework to handle this task was built under the name of Modern Synthesis [1]. It encompassed Darwin's idea of evolution by natural selection as an explanation for diversity and adaptation and

Mendel's idea of particular inheritance, giving rise to population and quantitative genetics, a theoretical frame that corroborated Darwin's hypothesis of the paramount power of selection for driving adaptive evolution [2]. This framework progressively aggregated multiple disciplines: behavioural ecology, microbiology, paleobiology, etc. Overall, this classic framework considers that the principal agency of evolution is natural selection of favourable variations, and that those variations are constituted by random mutations and recombination in a Mendelian population. The processes of microevolution, modelled by population and quantitative genetics, are likely to be extrapolated to macroevolution [3]. To this extent, models that focus on one or two loci are able to capture much of the evolutionary dynamics of an organism, even though in reality many interdependencies between thousands of loci (epistasis, dominance, etc.) occur as the basis of the production and functioning of a phenotypic trait. Among forces acting on populations and modelled by population geneticists, natural selection is the one that shapes traits as adaptations and the design of organisms; adaptive radiation then explains much of the diversity; and common descent from adapted organisms explains most of the commonalities across living forms (labelled homologies), and allows for classifying living beings into phylogenetic trees. Evolution is gradual because the effects of mutations are generally small, large ones being most likely to be deleterious as theorized by Fisher's geometric model [4].

Many theoretical divergences surround this core view: not everyone agrees that evolution is change in allele frequencies, or that population genetics captures the whole of the evolutionary process, or that the genotypic viewpoint — tracking the dynamics of genes as 'replicators' [5] or the strategy 'choices' of organisms as fitness maximizing agents [6] — should be favoured to understand evolution. Nevertheless, it has been a powerful enough framework to drive successful research programs on

* Correspondence: eric.bapteste@upmc.fr
[1]Sorbonne Universités, UPMC Université Paris 06, Institut de Biologie Paris-Seine (IBPS), F-75005 Paris, France
[2]CNRS, UMR7138, Institut de Biologie Paris-Seine, F-75005 Paris, France
Full list of author information is available at the end of the article

speciation, adaptation, phylogenies, evolution of sex, cooperation altruism, mutualism, etc., and incorporate apparent challenges such as neutral evolution [7], acknowledgement of constraints on variation [8], or the recent theoretical turn from genetics to genomics following the achievement of the Human Genome Program [9]. Causation is here overall conceived of as a linear causal relation of a twofold nature: from the genotype to the phenotype (assuming of course environmental parameters), and from the environment to the shaping of organisms via natural selection. For instance, in the classic case of evolution of peppered moths in urban forests at the time of the industrial revolution, trees became darkened with soot, and then natural selection favored darker morphs as 'fitter' ones, due to their being less easily detected by predator birds, resulting in a relative increase in frequency of the darker morphs in the population [10].

Yet in the last 15 years biologists and philosophers of biology have regularly questioned the genuinely unifying character of this Synthesis, as well as its explanatory accuracy [11]. Those criticisms questioned notably the set of objects privileged by the Modern Synthesis, arguably too gene-centered [12], and its key explanatory processes, since niche construction [13], lateral gene transfer [14, 15], phenotypic plasticity [16, 17], and mass extinction [18] could, for example, be added [11]. Usually these critiques emphasize aspects rooted in a particular biological discipline: lateral gene transfer from microbiology, plasticity from developmental biology, mass extinction from paleobiology, ecosystem engineering from functional ecology, etc. There were also recurring claims for novel transdisciplinary fields: evo-eco-devo [19], investigating the evolutionary dynamics of host and microbe associations (forming combinations often referred to as holobionts), evolutionary cell biology [20], or microbial endocrinology [21], among others. This latter discipline aims at understanding the evolved interactions between microbial signals and host development. Indeed, it is compelling for evolutionary biologists to decipher how such multi-species interactions became established (namely, whether they involved specific microbial species and molecules, and whether they evolved independently in different host lineages).

Evolutionary biology is thus currently undergoing various theoretical debates concerning the proper frame to formulate it [11, 22–24]. Here, we introduce an original solution which moves this debate forward, acknowledging that nothing on Earth evolves and makes sense in isolation, thereby challenging the key assumption of the Modern Synthesis framework that targeting the individual gene or organism (even when in principle knowing that it is part of a set of complex interactions) allows us to capture evolution in all its dimensions. Since the living world evolves as a dynamic network of interactions, we argue that evolutionary biology could become a science of evolving networks, which would allow biologists to explain organisational complexity, while providing a novel way to reframe and to unify evolutionary biology.

Biology is regulated by networks
Networks at the molecular level
Although numerous studies have focused on the functions of individual genes, proteins and other molecules, it is increasingly clear that each of these functions belongs to complex networks of interactions. Starting at the molecular scale, the importance of a diversity of molecular agents, such as (DNA-based) genes and their regulatory sequences, RNAs and proteins, is well recognized. Importantly, in terms of their origins and modes of evolution, these agents are diverse. Genes are replicated across generations, via the recruitment of bases along a DNA template, thereby forming continuous lineages, affected by Darwinian evolution. By contrast, proteins are reconstructed by recruitment of amino acids at the ribosomal machinery. There is no physical continuity between generations of proteins, and thus no possibility for these agents to directly accumulate beneficial mutations [25]. Moreover, all these molecular entities are compositionally complex, in the sense that they are made of inherited or reassembled parts. *E pluribus unum*: genes and proteins are (often) conglomerates of exons, introns [26–28], and domains [29–31]. Similar claims can be made about composite molecular systems, such as CRISPR and Casposons [32, 33], etc. This modular organisation has numerous consequences: among them, genes can be nested within genes [34]; proteins congregate in larger complexes [35]. Importantly, this modularity is not the mere result of a divergence from a single ancestral form, but also involves combinatorial processes and molecular tinkering of available genetic material [36–38]. The coupling and decoupling of molecular components can operate randomly, as in cases of presuppression proposed to neutrally lead to large molecular complexes [39–41]. Presuppression, also known as constructive neutralism, is a process that generates complexity by mechanically increasing dependencies between interacting molecules, in the absence of positive selection. When a deleterious mutation affects one molecular partner, existing properties of another molecule with which the mutated molecule already interacted can compensate for its partner defect. Presuppression operates like a ratchet, since the likelihood to restore the original independency between molecules (by reverting the deleterious mutation) is lower than the likelihood to move away from this original state (by accumulating other mutations). Molecular associations can also evolve under constraints [42], eventually reinforcing the relationships between molecular partners, as suggested for some operons [43] and fused genes [44, 45].

Consistently, interconnectedness is a striking feature of the molecular world [46, 47]. Genes belong to regulatory networks with feedback loops [48]. Proteins belong to protein–protein interaction networks. This systemic view contrasts with former atomistic views assigning one function to one gene. First, it is not always correct that a gene produces only a protein, in the case of alternative splicing. Second, it is also unlikely that a protein performs one function, because no protein acts alone. Rather, biological traits result from co-production processes. This is nicely illustrated by the actual process of translation, during which both proteins and DNA necessarily interact, allowing for the collective reproduction of these two types of molecular agents. How these different components became so tightly integrated is a central issue for explaining evolution. Understanding how the molecular world functions and evolves therefore requires analysing molecular organisation and the evolution of the architecture of interaction networks, especially since this structure can partly explain molecular reactions [46, 47, 49, 50]. Thus, systems biologists search for common motifs in molecular interaction networks from different organisms, such as feed-forward loops, assuming that some of these recurring patterns, because they affect different gene or protein sets, may reflect general rules and constraints affecting the construction and evolution of biological organisations [46].
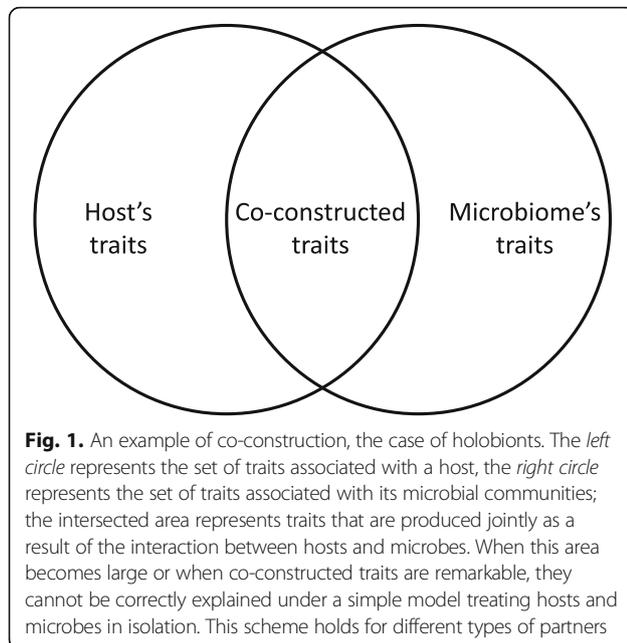
Focusing evolutionary explanations on the structure of the interactions between genes rather than on the primary sequence of the genes is fundamentally different from sequencing genes and inferring history from their sequences alone. One could think here of the case of explaining gene activation/repression. Comparative works on molecular interaction networks show that interactions affect the evolution of the molecules composing networks, which means that beyond compositional complexity, organisational complexity must be modeled to understand biological evolution [46, 51–54]. Before the analysis of complex networks, compensatory sets of elements, such as groups of sub-functional paralogous genes [55], or groups of genes with pressupressed mutations [39, 40], already stressed the evolutionary interdependence of molecules. However, compensatory interactions between agents, each of them being by themselves poorly adapted, ran counter to the intuition that natural selection will eliminate dysfunctional individual entities. Their recognition invites one to consider Earth as possibly populated by unions of individually dysfunctional agents rather than by the fittest survivors within individual lineages, possibly since early life, according to Woese's theory on progenotes, namely communities of interacting protocells unable to sustain themselves alone, evolving via massive lateral genetic exchanges [56].

At the molecular level, it is reasonable to assume that processes resulting from interactions of a diversity of intertwined agents offer a crucial *explanans* of biological complexity. Rather than 'one agent, one action', it would be more accurate to consider 'a relationship between agents, one action' as the *modus operandi* of life. Multiple drivers, of different nature, contribute to the evolution of these interactions: among others, gene co-expression/co-regulation [57], sometimes mediated by transposons [58–61]; the evolutionary origin of the genes [62]; and also physical and chemical laws, as well as the presence of targeting machineries that constrain and regulate diffusion processes in the cell. These types of relationships described at the molecular level are also recovered at other levels of biological organisations.

Networks at the cellular level

Similar conclusions have been reached at the cellular level, also crucial for understanding life history. All prokaryotes and protists are unicellular organisations, and the cell is a fundamental building block of multicellular organisms. Cells must constantly evaluate the states of their inner and outer environments, i.e. to adjust their gene expression and react accordingly [46]. This involves regulatory, transduction, developmental, and protein interaction networks, etc. Cells are built upon inner networks of interacting components, and involved in or affected by a diversity of exchanges, influences and modes of communications (namely, genetic, energetic, chemical and electrical modes). Microbiology has gone a long way toward unraveling these processes since its heyday of pure culture studies, a fruitful reductionist approach now complemented by environmental studies. These latter further unraveled that cells compete and cooperate with, and even compensate for each other, within mono- or multispecific microbiomes [63, 64]. Both types of microbiomes have a fundamental commonality: they produce collective properties and co-constructed phenotypes (Fig. 1) evolving at the interface between cells. Such properties cannot be understood without considering networks of influences: the oscillatory growth of biofilms of *Bacillus subtilis* cannot be deduced from the analyses of the complete genomes of these clones, but requires modeling metabolic co-dependence within a monogenic community affected by a delayed feedback loop, involving chemical and electrical signals [65, 66].

Furthermore, many cellular agents show a relative lack of autonomy. In nature, some groups of prokaryotes display complementary genomes with incomplete metabolic pathways, consistent with the black queen hypothesis, which predicts that our planet is populated by groups of (inter)dependent microbes [67, 68]. More precisely, this hypothesis predicts the loss of a costly function, encoded by a gene or a set of genes, in individuals, when this

Fig. 1. An example of co-construction, the case of holobionts. The *left circle* represents the set of traits associated with a host, the *right circle* represents the set of traits associated with its microbial communities; the intersected area represents traits that are produced jointly as a result of the interaction between hosts and microbes. When this area becomes large or when co-constructed traits are remarkable, they cannot be correctly explained under a simple model treating hosts and microbes in isolation. This scheme holds for different types of partners

function becomes dispensable at the individual level, since it is achieved by other individuals that produce (usually leaky) public goods in sufficient amount to support the equilibrium of the community. Thus, gene losses in some cells are compensated by leaks of substrates from other cells, formerly encoded by the lost genes. Some microbes experience labor division [69]. Symbionts and endosymbionts depend on their hosts. The 'kill the winner' theory [70] further challenges the notion that the microbial world is a world of fit cellular individuals. This theory stresses a collective process via which viruses mechanically mostly attack cells that reproduce faster and thus regulate bacterial populations, these latter sustaining their diversity because these populations are comprised of individual prokaryotic cells that make a suboptimal use of a diversity of resources. Thus, cells belong to networks that affect their growth and survival, which might explain why most bacteria cannot be grown in pure culture. They only truly thrive within communities, whose global genetic instructions are spread over several genetically incomplete microbes.

Accounting for these internal and external cellular networks requires considering processes that are not central in the synthetic evolutionary theory. Typically, the notion that cellular evolution makes jumps, because new components and processes (such as metabolic pathways) are acquired from outside a given cellular lineage, contrasts with more gradual accounts of biological change, like accounts based on point mutations affecting genes already present in the lineage. Because saltations (macromutations) are essential evolutionary outcomes of introgressive processes, via the combination

of components from different lineages, no complete picture of evolution can be provided without these jumps, which are naturally modeled by networks. Indeed, genetic information has been flowing both vertically and horizontally between prokaryotes for over 3.5 billion years [71–77], and possibly earlier, according to Woese, who proposed that our universal ancestor was not an entity but a process, that is, genetic and energetic exchanges within protocellular communities [56]. Remarkably, this latter case indicates that network modeling could help to tackle a fundamental issue in evolutionary biology: modeling the evolution of biological processes that emerge from interactions between biological entities. Since these interactions can be represented by a network, the evolution of these interactions, describing the evolution of biological processes, can then be represented by dynamic networks. Likewise, eukaryogenesis rested on the co-construction of a novel type of cell, as a result of the endosymbiosis of a bacteria within an archaeon [78–80]. Later, the evolution of photosynthetic protists emerged from endosymbioses involving unicellular eukaryotes and cyanobacteria, or various lineages of protists, namely in secondary and tertiary endosymbioses [81]. Such endosymbioses, and their outcomes as illustrated in our work [82, 83], are also naturally modeled using networks.

Moreover, the long-term impact of these introgressive processes on cellular evolution should not be underestimated. For instance, endosymbiosis does not merely introduce new cellular lineages, it also favors the evolution of chimeric structures and chimeric processes within cells [83–91]. Such intertwining cannot be modeled using a single genealogical tree, which only recapitulates cellular divergence from a last common ancestor. Even though cells always derive from other cells, a full cellular history cannot be reduced to the history of some cellular components that are assumed to track the history of cellular division [92]. In particular, phylogenetic analyses of informational genes cannot be the only clue to understanding the origins of cellular diversity, since these genes do not reflect how cells are organized, how they gather their energy, and how they interact with each other. Analyzing the co-construction side of evolution requires enhanced models: understanding eukaryotic evolution requires mixed considerations of cellular architecture, population genetics and energetics, which go beyond classic phylogenetic models, which not so long ago were still prone to considering three primary domains of life [93–95].

Although invoking multiple agents rather than a single ancestor in evolutionary explanations might appear to contradict the famous Ockham's razor [96], it does so only superficially when it is likely that many cells are co-constructed, especially in the context of a web of life. Enhanced models including intra- and extracellular

interactions appear necessary to understand cellular complexity, including the predictable disappearance of traits (and processes), namely the convergent gene loss of mitochondria and plastids [97] by a process called dedarwinification [98, 99].

Networks beyond the cellular level

Studies of multicellular organisms—we will focus on animals—have led to similar general findings. Understanding animal traits and their evolution requires analyzing the relationships between a multiplicity of agents belonging to different levels of biological organisation, eventually nested, some of which co-constructs animals and guarantees their complete lifecycle [100]. Because no sterile organism lives on Earth, animal development, health and survival depend on microbes. Granted, bacteria can often legitimately be seen as part of the environmental demands in an evolutionary model focused on the host's lineage; or sometimes bacteria and host could also be considered as part of a coevolution process, with no need to posit the whole as a unit of selection [101]. However, asking 'who is the beneficiary of the symbiosis as the result of evolution?' may in some cases lead to the recognition that bacteria and host evolved together and were selected together [102]. More generally, while some microbes contribute to animals' lives possibly as a result of host-derived selection, others contribute as a result of selectively neutral processes (like microbial priming [103]) [101, 104]. These interactions produce communication networks within the animal body: chemical information circulates between the animal brain and the gut microbiome. These interactions also result in communication and interaction networks between individuals. In some animal lineages, the microbiome affects social behaviors, for instance fermenting microbes inform about the gender and reproductive status in hyena [105]. Components of the microbiome also affect mating choice [106], reproductive isolation and possibly speciation. Consequently, the microbiome now appears as an essential component of animal studies [107]. Microbiome studies, the significance of which is overstated in some respects, nevertheless have shown that the evolutionary intertwining between many metazoa and commensal or symbiotic bacteria could not be neglected anymore and black-boxed in favor of purely host gene-centered evolutionary models. And the associations between hosts and microbes do not need to be units of selection to be part of the recent insights that support the novel theoretical framework proposed here. Their interplay imposes reconfigurations of practices, theories and disciplines [108]. As a result of our improved insight into evolution, zoology and immunology [109] become theaters of new ecological considerations [110], sometimes strangely qualified as Lamarckian [111, 112], because animals can

recruit environmental microbes and transmit them (with a non-null heritability [113]) to their progeny. Therefore, nuclear gene inheritance alone may provide too narrow a perspective to account for the evolution of all animal traits; as an example, aphid body color depends on animal genetics and the presence of *Rickettsiella* [114]. Population genetics gets included in a broader community genetics, which also considers transmission of microbes and their genes [108, 114]. The use of gnotobiotic and transbiotic animals becomes a new experimental standard to analyze multigenomic collectives without counterparts in modern synthesis theories. These collectives harbor morphological, physiological, developmental, ecological, behavioral and evolutionary features [115–119] that are not purely constructed by animal genes, but rather appear to be co-constructed at the genetic and metabolic interface between the microbial and macrobial worlds, while the content of the respective animal genomes only provides incomplete instructions. Understanding animal evolution requires understanding the interaction networks between components from which these taxa evolved, and the networks to which these taxa still belong.

In ecology, an analogous turn towards network thinking has been promoted since the 1990s with the general acceptance of the notions of metapopulations [120] and then metacommunities [121]. These views suggest that the dynamics of ecological biodiversity is not so much located within a community of species but rather in a metacommunity, which can be thought of as a network of communities exchanging species, while targeting one community blinds one to what genuinely accounts for biodiversity and ecosystem functioning [122].

This quick overview provides evidence that networks are at the origin of the genes of unicellular and multicellular organisms and central for their functions. The living world is a world of 'and' and 'co-'. From division of labor and compensations, to dependencies and co-constructions, etc.: interactions (which only begin to be deciphered) are everywhere in biology. Thus, explaining the actual features of biodiversity requires explaining how multiple processes, interface phenomena (co-construction of biological features, niche construction, metabolic cooperation, co-infection and co-evolution) and organisations (for instance, from molecular pathways to organisms and ecosystems) arose from interacting components, and how these processes, phenomena and organisations may have been sustained and transformed on Earth.

Reframing evolutionary explanations from the scaffolded evolution perspective

Introducing a classification of interacting components

While classic evolutionary models, prompted by Darwin's famous tree [123], mostly stress how related

entities diverge in relative independence, it appears important to show how a diversity of components, which may not be related, interact and produce various evolutionary patterns.

The notion of scaffolding [124], which describes how one entity continues an event initiated by another entity, and relies on it up to the point that at some timescale it becomes dependent upon it for further evolution, appears as a fundamental relationship to describe the evolution of life. We propose scaffolding should become more central in explanations of evolution because no components from the biological world are actually able to reproduce, or persist, alone (Fig. 2). Each entity influences or is influenced by something external to it, and is consequently part of a process. Scaffolding thus defines the causal backbone of collective evolution. It describes the historical continuity between temporal slices of interaction networks, since any evolutionary stage relies on previously achieved networks and organisations. Therefore, describing the evolution of interactions requires explanations to address the following issues: what scaffolds what, what transforms the environment of what, and are these influences reciprocal? Characterizing the types of components that, together, have evolutionary importance through their potential interaction is therefore a central step to expanding evolutionary theory.

We propose that a first distinction can be made between obligate and facultative components. Suppressing the former impacts the course and eventually the reproduction of the process to which they contribute (Fig. 3), whereas facultative components do not hold such a crucial role, and may simply be involved by chance. A second distinction is whether the components are biotic (genes, proteins, organisms...) or abiotic (such as minerals, environmental, cultural artefacts). Abiotic components can be recruited from the environment or be shaped by biological processes [125]. They can also alter the evolution of the biotic components, for example, environmental change can drive genetic and organismal evolution and selection. The history of life clearly depends on the interplay of both types of components. Biotic components, however, deserve a specific focus. Some of them form lineages (for instance, genes replicate), while others do not (for instance, proteins are reconstructed). Finally, interacting replicated components can be further classified into fraternal components when they share a close last common ancestor (e.g. in kin selection cases), and egalitarian components, when they belong to distinct lineages (as an example, think of the evolution of chimeric genes by fusion and shuffling [29, 45, 126]) [63].

Introducing dynamic interaction networks

Biodiversity usually evolves from interactions between the diverse types of components described above. For example, metalloproteases emerge from the interaction between reconstructed biotic components (proteins) and a metal ion. Regulatory networks involve biotic components that can be either replicated (i.e. genes and promoters) or reconstructed (i.e. proteins). Protein interaction networks intertwine reconstructed egalitarian biotic components, which means proteins that are not homologous. Evolutionary transitions such as eukaryogenesis result from the interweaving of biotic components (cells) from multiple lineages. Holobionts evolve from interactions between egalitarian biotic components (macrobial hosts and microbial communities) and possibly abiotic components, such as the mineral termite mounds, or the volatile chemicals produced by the microbial communities of hyenas [105].

Taking collectives of interacting components as central objects of study in evolutionary biology invites us to expand the methods of this field. It encourages developing statistical approaches or inference methods beyond those
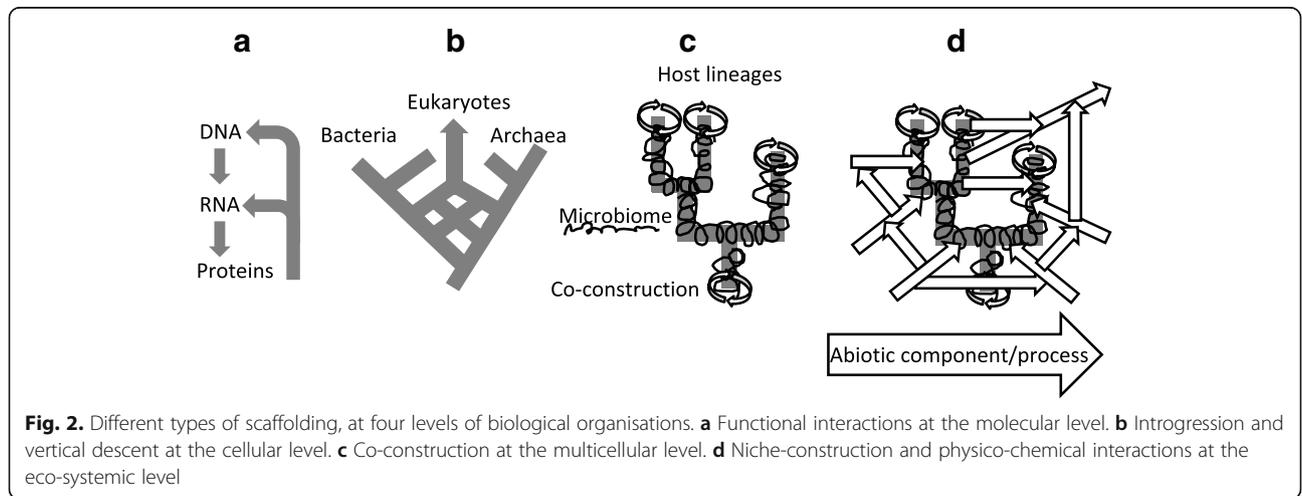

Fig. 2. Different types of scaffolding, at four levels of biological organisations. **a** Functional interactions at the molecular level. **b** Introgression and vertical descent at the cellular level. **c** Co-construction at the multicellular level. **d** Niche-construction and physico-chemical interactions at the eco-systemic level
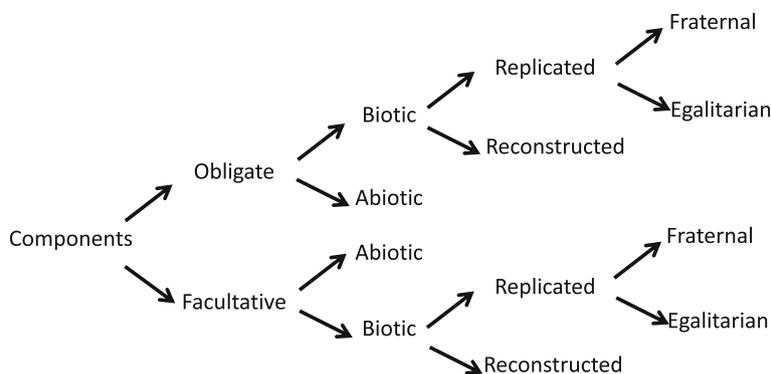
Fig. 3. Classification of major types of components in evolving systems. A process/collective cannot be completed in the absence of obligate components, whereas facultative components do not affect the outcome of the process/function of the collective. Biotic components are biological, material products, whereas abiotic components are environmental, geological, chemical, physical or cultural artefacts. Replicated components are produced by replication, which implies a physical continuity between ancestral and descendent components; they undergo a paradigmatic Darwinian evolution. Reconstructed components are reproduced without direct physical continuity, and cannot directly accumulate beneficial mutations. Fraternal components belong to the same lineage, whereas egalitarian components belong to different lineages

operating under the very common assumption that biological components are independent. Therefore, we propose to represent interactions between components in the form of networks in which components are nodes and their interactions (of various sorts) are edges. These networks are conceptually simple objects. They can be described as adjacency lists of interactions, in the form 'component A interacts with component B, at time t (when such a temporal precision is known)'. Such dynamic interaction networks could become more central representations and analytical frameworks, and serve as a common *explanans* for various disciplines in an expanded evolutionary theory. Importantly, because these networks embed both abiotic and biotic, related and unrelated components (like viruses, cells and rocks), they should not be conflated with phylogenetic networks, but

recognized as a more inclusive object of study (Fig. 4). Where phylogenies describe relationships, networks can describe organisations. How such organisations evolve could for example be described by identifying evolutionary stages, that is, sets of components and of their interactions simultaneously present in the network (Fig. 4). Investigating the evolution of an ecosystem corresponds to studying the succession of evolutionary stages in such networks and detecting possible regularities—in the sense that some evolutionary stages would fully or partly reiterate over time—or hinting at rules or constraints (like architectural contingencies [127, 128] or principles of organisations [46]) on the recruitment, reproduction and heritability of their components.

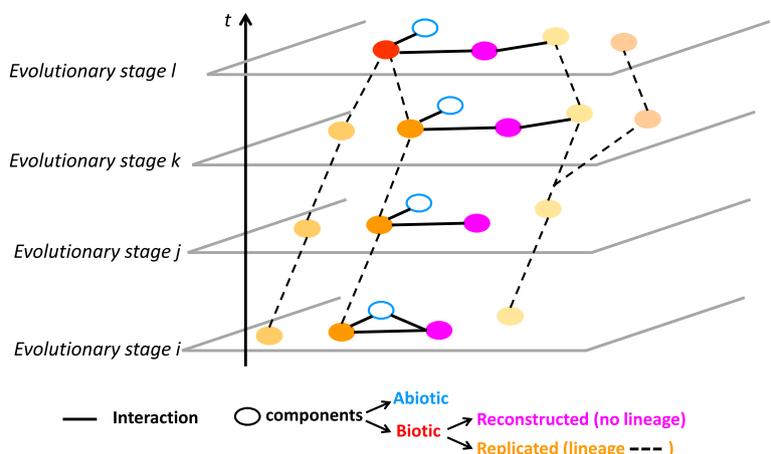Thus, we suggest that evolutionary biology could be reframed as a science of evolving networks, because

Fig. 4. An evolving interaction network. Nodes are components (*circles* are full when the component is biotic). *Thick black edges* represent interactions between these components. The network topology evolves as nodes or their connection change. *Dashed edges* represent the phylogenetic ancestry of lineage-forming components

such a shift would allow inclusive, multilevel studies of a larger body of biological and abiotic data, via approaches from network sciences.

Concrete strategies to enhance network-based evolutionary analyses

Enhancing network-based evolutionary analyses, beyond the now classic research program of phylogenetic networks, could consolidate comparative analyses in the nascent field of evolutionary systems biology [129, 130], as illustrated by examples based on molecular networks. Network construction/gathering constitutes the first step of such analyses. This involves first defining nodes of the network, namely components suspected to be involved in a given system, and edges, namely qualitative (or quantitative, when weighted) interactions between these entities. Many biological interaction networks (gene co-expression networks (GCNs), gene regulatory networks (GRNs), metabolic networks, protein–protein interaction networks (PPIs), etc. [46]) are already known for some species, or can be inferred [131–136]. For example, GCNs offer an increasingly popular resource to study the evolution of biological pathways [137], as well as to reveal conservation and divergence in gene regulation [138]. GCNs are already used for micro-evolution studies, as in the case of fine-grained comparisons of expression variations between orthologous genes across closely related species, and for the analysis of minor evolutionary and ecological transitions, such as changes of ploidy [139, 140], adaptation to salty environments [141] or drugs [142], or the effects of plant domestication [143, 144]. Likewise, GRNs are starting to be used in micro-evolution and phenotypic plasticity studies [145]. Understanding the dynamics of GRNs appears critical to inferring the evolution of organismal traits, in particular during metazoan [146–148], plant [149] and fungal [150] evolution. We suggest that PPI, GCN and GRN studies could become mainstream and also be conducted at (much) larger evolutionary and temporal scales, to analyze additional, major, transitions.

Based on these established networks, two major types of evolutionary analyses (network-decomposition and graph-matching; Fig. 5) can be easily further developed by evolutionary biologists. More precisely, the above-mentioned kinds of biological networks could be systematically turned into what we call evolutionary colored
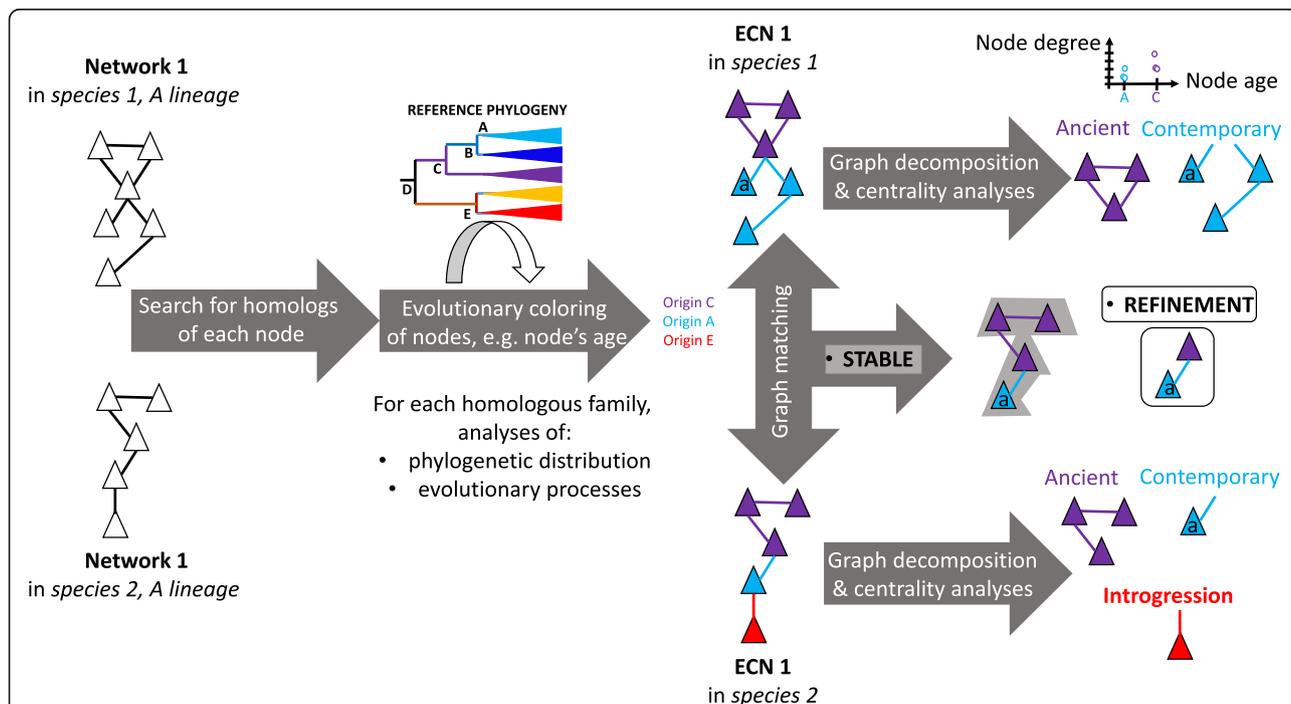

Fig. 5. Workflow of the evolutionary analysis of interaction networks. From *left* to *right*: *triangles* represent components of interaction networks, *edges* between *triangles* represent interactions between these components. Interaction networks are first constructed/inferred, then their nodes and edges are colored to produce evolutionary colored networks (ECNs) that represent both the topological and the evolutionary properties of the networks. ECNs can be investigated individually by graph decomposition and centrality analyses, or several ECNs can be compared by graph alignment. The two types of comparisons can return conserved subgraphs that allow understanding of the dynamics of interaction networks, meaning when different sets of interactions (hence processes) evolved, and whether these interactions were evolutionarily stable. *Ancient* and *Contemporary* refer to the relative age of the sub-graphs, identifying new clade-specific relationships (here called refinement); introgression indicates that a component, and the relationship it entertains with the rest of the network, was inferred to result from a lateral acquisition

biological networks (ECNs). In ECNs, each node of a given biological network is colored to reflect one or several evolutionary properties. For example, in molecular networks, nodes correspond to molecular sequences (genes, RNA, proteins) that belong to homologous families that phylogenetic distribution across host species allows us to date [137, 151–156]. The 'age' of the family at the node can thus become one evolutionary color (Fig. 5). Likewise, several processes affecting the evolution of a molecular family (selection, duplication, transfer, and divergence in primary sequence) can be inferred by classic phylogenetic analyses or, as we proposed, by analyses of sequence similarity networks [157]. Such studies provide additional evolutionary colors (like quantitative measures: intensity of selection, rates of duplication, transfer, and percentage of divergence), which can be associated with nodes in ECNs [139, 149, 154, 158–161]. Thus, ECNs contain both topological information, characteristic of the biological network under investigation, as well as evolutionary information: what node belongs to a family prone to duplication, divergence, or lateral transfer, as well as when this family arose. Combining these two types of information in a single graph allows us to test specific hypotheses regarding evolution.

Using ECNs, it is first fruitful to test whether (or which of) these evolutionary colors correlates with topological properties of the ECNs [162–164]. The null hypothesis that nodes' centrality, e.g. nodes' positions in the network, is neither correlated with the age nor with the duplicability, transferability or divergence of the molecular entities represented by these nodes can be tested. Rejection of this hypothesis would hint at processes that affect the topology of biological networks or are affected by the network topology. For example, considering degree in networks, proteins with more neighbors are less easily transferred [163], highly expressed genes, more connected in GCNs, evolve slower than weakly expressed genes [165], and genes with lower degrees have higher duplicability in yeast, worm and flies [166]. Considering position in networks, node centrality correlates with evolutionary conservation [136], gene eccentricity correlates with level of gene expression and dispensability [167], and proteins interacting with the external environment have higher average duplicability than proteins localized within intracellular compartments [168]. Additionally, network structure gives a clue to evolution since old proteins have more interactions than new ones [169, 170]. Generalizing these disparate studies could help to understand the dynamics of biological networks, in other words how the architecture, the nodes and edges of present day networks, evolved and whether their changes involved random or biased sets of nodes and edges or follow general models of network growth with detectable drivers.

This focus would complement a classic tree-based view. For instance, under the reasonable working hypothesis that pairs of connected nodes of a given age reflect an interaction between nodes that may have arisen at that time [154, 171], ECNs can easily be easily decomposed into sub-networks, featuring processes of different ages (that is, sets of nodes of a given age, e.g. sets of interacting genes). This strategy allows identification of conserved network patterns, possibly under strong selective pressure [159]. Constructing and exploiting ECNs from bacteria, archaea, and eukaryotes thus has the potential to define conserved ancestral sets of relationships between components, allowing evolutionary biologists to infer aspects of the early biological networks of the last common ancestor of eukaryotes, archaea and bacteria and even of the last universal common ancestor of cells. Assuming that some of these topological units correspond to functional units [172], especially for broadly conserved subgraphs [138, 149, 152, 166, 173–182], would allow network decompositions to propose sets of important processes associated with the emergence of major lineages.

Moreover, graph-matching of ECNs allows several complementary analyses. First, for interaction networks, such as GRNs, whose sets of components and edges evolve rapidly [183–185], it becomes relevant to analyze where in the network such changes occur in addition to (simply) tracking conserved sets of components and edges. Whereas the latter can test to what extent conservation of the interaction networks across higher taxa supports generalizations made from a limited number of model species [186], the former allows us to test a general hypothesis: are there repeated types of network changes? For example, does network modification primarily affect nodes with particular centralities, as exemplified by terminal processes [187], or modules? Systematizing these analyses would provide new insights into whether the organisation principles of biological networks changed when major lineages evolved or remained conserved. In terms of the ECN, can the same model of graph evolution explain the topology of ECNs from different lineages? The null hypothesis would be that these major transitions left no common traces in biological networks. An alternative hypothesis would be that the biological networks convergently became more complex (more connected and larger) during these transitions to novel life forms. Indeed, analyses conducted on a few taxa have reported quantifiable and qualifiable modifications in biological networks (in response to environmental challenges [188], during ecological transitions [189] or as niche specific adaptations [190]). More systematic graph-matching [191–193] and motif analyses, comparing the topology of ECNs from multiple species, could likewise be used to test the hypothesis

that major lineages are enriched in particular motifs (either modules of colored nodes and edges, or specific topological features, such as feed-forward loops [46] or bow-ties [194]). It would also allow identification of functionally equivalent components across species, namely different genes with similar neighbors in different species [176].

While inferences on conserved sets of nodes and edges in ECNs are likely to be robust (since the patterns are observed in multiple species), missing data (missing nodes and edges) constitute a recognized challenge, especially for the interpretation of what will appear in ECN studies as the most versatile (least conserved) parts of the biological networks. The issue of missing data, however, is not specific to network-based evolutionary analyses, and should be tackled, as with other comparative approaches, by the development and testing of imputation methods [195–197]. Moreover, issues of missing data can also be addressed by the production of high coverage -omics datasets in simple systems, allowing for (nearly) exhaustive representations of the entities and their interactions (i.e. PPIs, GCNs and GRNs within a cell, or metabolic networks within a species poor ecosystem). This kind of data would allow testing for the existence of selected emergent ecosystemic properties (like carbon fixation), as stated by the ITSNTS hypothesis [198]. For instance, deep coverage time series of metagenomic/metatranscriptomic data coupled with environmental measures from a simple microbial ecosystem, such as carbon fixation, could produce enough data to allow the evolutionary coloring of nodes of metabolic networks. Comparing ECNs representing, at each time point, the origin and abundance of the lineages hosting the enzymes involved in carbon fixation could test whether some combinations of lineages are repeated over time, and whether the components (e.g. genes and lineages) vary, whereas carbon fixation is maintained in the ecosystem, which would suggest that this process evolves irrespective of the nature of the interacting components.

Finally, entities from different levels of biological organisation (domains, genes, genomes, lineages, etc.) could also be studied together in a single network framework, by integrating them into multipartite networks [199]. Recently, our studies and others (see [200] and references therein) have demonstrated that various patterns in multipartite graphs can be used to detect and test combinatorial (introgressive) and gradual evolution (by vertical descent) affecting genes and genomes. Decomposing multipartite networks into twins and articulation points could for example then be used to represent and analyze the evolution of complex composite molecular systems, such as CRISPR, or the dynamics of invasions of hairpins in genomes [201].

Further justifications for a shift toward network thinking

Enlargement of evolutionary biology

Focusing evolutionary explanations and theories on collectives of interacting components, which may be under selection, facilitate selection, or condition arrangements through neutral processes [39, 40, 202], and representing these scaffolding relationships using networks with biotic and abiotic components and a diversity of edges representing a diversity of interaction types would be an enlargement. Enlargements, as expressing the need to consider structures that are more general than what already exists, have already occurred within evolutionary theory, when simplifications from population genetics were relaxed with respect to the original formalization in the Modern Synthesis [203], to account for within-genome interaction [9], gene–environment covariance [204], parental effects [205], and extended fitness though generations [206]. It also occurred when reticulations representing introgressions were added to the evolutionary tree.

Interestingly, replacing standard linear models in evolutionary theory with network approaches would transcend several traditional axes structuring the debates in evolutionary biology. For instance, scaffolded evolution, the idea that evolution relies on what came before, is orthogonal to the distinction between vertical and horizontal descent, since both tree-like and introgressive evolution are particular cases of scaffolding. Scaffolded evolution is also orthogonal to the distinction between gradual and saltational evolution. Likewise, scaffolded evolution is orthogonal to the debates between the actual role of adaptations vs neutral processes. Selection is a key mode of evolution of collectives but not the only one. The processes involved in the forming and evolution of collectives are not even restricted to the key processes of the Modern Synthesis (drift, selection, mutation and migration) but embrace interactions such as facilitation—namely antagonistic interactions between two species that allow a third species to prosper by restraining one of its predators or parasites [207], presuppression [39, 40], etc. Consequently, some evolutionary concepts may become more important than they currently are to explain evolution. For example, contingency, which means the dependence of an evolutionary chain of events upon an event that itself is contingent, in the sense that it can't be understood as a selective response to environmental changes [18, 208, 209], is often associated with extraordinary events, like mass decimation. Contingency could come to be seen as a less extraordinary mode of evolution in the history of life, since the ordinary course of evolution might include many cases of contingent events, that is, associations of entities in a transient collective, including any scaffolds—associations that are not necessarily selective responses or the outcomes of processes modeled in population genetics.

Likewise, adopting a broader ontology could affect how evolutionary theorists think about evolution. Population thinking and tree-thinking came after essentialist conceptions of the living words, when populations and lineages were recognized as central objects of evolutionary studies [210]. A shift towards collectives and scaffolded evolution might encourage a similar development: the emergence of an openly pluralistic processual thinking, consistent with Carl Woese's proposal to reformulate our view of evolution in terms of complex dynamic systems [211].

Further unifying the evolutionary theory

Using a network-based approach to analyse dynamic systems also permits explanations that rely purely on statistical properties [212] or on topological or graph theoretical properties [213, 214] besides standard explanations devoted to unravelling mechanisms responsible for a phenomenon. Moreover, because of the inclusiveness of the network model, disciplines already recognized for their contribution to evolutionary theory (microbiology, ecology, cell biology, genetics, etc.) could become even more part of an interdisciplinary research program on evolution, effectively addressing current issues, consistent with the repeated calls for transdisciplinary collaborations [19–21, 215]. Disciplines that were not central in the Modern Synthesis—chemistry, physics, geology, oceanography, cybernetics or linguistics—could aggregate with evolutionary biology. Since a diversity of components gets connected by a diversity of edges in networks featuring collectives, as a result of a diversity of drivers, several explanatory strategies could be combined to analyze evolution. This extension to seemingly foreign fields makes sense when the components/processes studied by these other disciplines are evolutionarily or functionally related to biotic components and processes (either as putative ancestors of biological components and processes, like the use of a proton gradient in cells, which possibly derived from geological processes affecting early life [216], or as descendants of biological systems, e.g. technically synthesized life forms, which have a potential to alter the future course of standard biological evolution).

Remarkably, this mode of unification of diverse scientific disciplines would be original: the integration would not be a unification in the sense of logical positivism [217]—namely reducing a theory to a theory with more basic laws, or a theory with a larger extension. It would be a piecemeal [218] unification. Some aspects would be unified through a specific kind of graph modeling (because some interactions, namely mechanical, chemical, ecological ones, and a range of time scales are privileged in a set of theories), while other theories might be unified by other graph properties (like different types of edges and components). For example, the fermentation hypothesis for mammalian chemical communication could be analyzed in a multipartite network framework, which would involve nodes corresponding to individual mammals, nodes corresponding to microbes, and nodes corresponding to odorous metabolites. Nodes corresponding to mammals could either be colored to reflect an individual's properties (its lineage, social position, gender, sexual availability), or these nodes could be connected by edges that reflect these shared properties, which defines a first host subnetwork. This host subnetwork can itself be further connected to a second subnetwork, namely the microbial subnetwork in which nodes representing microbes, colored by phylogenetic origins, could be connected to reflect microbial interactions (gene transfer, competition, metabolic cooperation, etc.). Connections between the host and microbial subnetworks could simply be made by drawing edges between nodes representing individual mammals hosting microbes, and nodes representing these microbes. Moreover, nodes representing mammals and nodes representing microbes could be connected to nodes representing odorous metabolites to show what odours are associated with what combinations of hosts and microbes. Elaborating this network in a piecemeal fashion would involve cooperation between chemists, microbiologists, zoologists and evolutionary biologists.

Of note, the use of integrated networks could pragmatically address a deep concern for evolutionary studies, by connecting phenomena that occur at different timescales: development and evolution [219] or ecology and evolution [220]. Considering transient collectives (thus processes) as stable entities at a given time-scale, when these collectives change much more slowly than the process in which they take part, amounts to a focus on interactions occurring at a given time scale by treating the slower dynamics as stable edges/nodes. Then, various parts of the networks embody distinct timescales, which may provide a new form of timescale integration, working out the merging of timescales from the viewpoint of the model, and with resources intrinsic to the model itself. The reason for this is that a node in an interaction network Ni, describing processes relevant at a time scale i, can itself be seen as the outcome of another (embedded) interaction network Nj, unfolding at a time scale j. This nestedness typically occurs when the node in Ni represents a collective process, involving components that evolve sufficiently slowly with respect to the system considered at the time scale i to figure as an entity, a node in Ni. In the case of a PPI network Ni, each node conventionally represents a protein, but the evolution of each protein could be further analysed as the result of mutation, duplication, fusion and shuffling events affecting the gene family coding the proteins over

time; for instance, each protein could thus be represented as the outcome of interaction between domains in a domain–domain interaction network Nj. Considering these two time-scales, it becomes apparent that gene families enriched in exon shuffling events, a process directly analysable in Nj, have a higher degree in PPI networks represented at the time-scale Ni [221].

Predictions: discovery of co-constructed phenotypes

What possible findings may result from this perspective shift? One can only speculate, but the nature of the potential discoveries is exciting. At the molecular level, the structure and composition of regulatory networks and protein interaction networks could be substantially enhanced to scaffolding elements. Currently, these networks represent interactions within a single individual/ species. Yet, viruses are everywhere, viral genes and proteins clearly influence the networks of their hosts, and likely constitute an actual part of their evolution. Thus, virogenetics, a novel transdiscipline, may prosper in an expanded evolutionary theory to show how and to what extent viruses co-construct their hosts, including perhaps reproductive-viruses, allowing their hosts to complete their lifecycles. At the cellular level, new modes of communication [222, 223] could be discovered, as possible viral and microbial languages and communication networks in biofilms would exemplify. At the level of multicellular organisms and holobionts, 'symbiotic codes', guiding the preferential association between hosts and symbionts, could be identified. At the level of phyla, hidden evolutionary transitions may be unraveled. While secondary (and tertiary) acquisitions of plastids have been documented [81], it might be shown that mitochondria too have been so acquired in some eukaryotic lineages (alongside the plastid or independently). Secondarily acquired mitochondria may provide their new hosts with additional compartments, where chimeric proteomes could assemble [91, 224] and perform original physiological processes. At the ecosystemic level, evolving networks could be used to model the changes and stases of our planet, grounding biotic lineages and processes in their environment, while highlighting potential regularities in the organisations and dynamics of ecosystems. What affects the stability of what over the course of evolution could thus become a central theme of an expanded evolutionary theory.

Concluding remarks and open questions

Interactions are not merely a part of biological history, they are what made this history. But evolutionary biologists have certainly not reconstructed the Dynamic Interaction Network of Life (DINoL) yet. Undertaking this endeavor, however, would emphasize the importance of processes. Our ancestors were processes. Our descendants and those of other life forms will be processes too. Some one hundred and fifty years after *On the Origin of Species*, which started a great evolutionary inquiry, evolutionists should prepare to face a larger challenge: expanding evolutionary theory to study the evolution of processes. With the development of -omics and network sciences, the concepts, data and tools for this research program are increasingly available.

Authors' contributions
EB and PH wrote, read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Sorbonne Universités, UPMC Université Paris 06, Institut de Biologie Paris-Seine (IBPS), F-75005 Paris, France. [2]CNRS, UMR7138, Institut de Biologie Paris-Seine, F-75005 Paris, France. [3]Institut d'Histoire et de Philosophie des Sciences et des Techniques (CNRS / Paris I Sorbonne), F-75006 Paris, France.

Published online: 29 May 2018

References
1. Huxley J. Evolution: the modern synthesis. Princeton: Princeton University Press; 1942.
2. Gayon J. Darwinism's struggle for survival: heredity and the hypothesis of natural selection. Cambridge: Cambridge University Press; 1998.
3. Simpson GG. Tempo and mode in evolution. New York: Columbia University Press; 1944.
4. Martin G, Lenormand T. The distribution of beneficial and fixed mutation fitness effects close to an optimum. Genetics. 2008;179:907–16.
5. Dawkins R. The extended phenotype. New-York: Oxford University Press; 1982.
6. Grafen A. A first formal link between the Price equation and an optimisation program. J Theor Biol. 2002;217:75–91.
7. Kimura M. The neutral theory of molecular evolution. Cambridge: Cambridge University Press; 1983.
8. Maynard Smith J, Burian R, Kauffman S, Alberch P, Campbell J, Goodwin B, et al. Developmental constraints and evolution. Q Rev Biol. 1985;60:265–87.
9. Griffiths P, Stotz S. Genetics and philosophy: an introduction. Cambridge: Cambridge University Press; 2013.
10. Kettlewell HDB. Selection experiments on industrial melanism in the Lepidoptera. Heredity. 1955;9:323–42.
11. Laland K, Uller T, Feldman M, Sterelny K, Muller GB, Moczek A, et al. Does evolutionary theory need a rethink? Nature. 2014;514:161–4.
12. Bateson P. The return of the whole organism. J Biosci. 2005;30:31–9.
13. Odling-Smee J, Laland K, Feldman M. Niche construction: the neglected process in evolution. Princeton: Princeton University Press; 2003.
14. Doolittle WF, Bapteste E. Pattern pluralism and the Tree of Life hypothesis. Proc Natl Acad Sci U S A. 2007;104:2043–9.
15. Sapp J. The new foundations of evolution: on the Tree of Life. New-York: Oxford University Press; 2009.

16. Walsh DM. Organisms, agency, and evolution. Cambridge: Cambridge University Press; 2015.

17. West-Eberhard MJ. Developmental plasticity and evolution. Oxford: Oxford University Press; 2003.

18. Gould SJ. Wonderful life. The Burgess shale and the nature of history. New York: Norton; 1989.

19. Gilbert SF, Bosch TC, Ledon-Rettig C. Eco-evo-devo: developmental symbiosis and developmental plasticity as evolutionary agents. Nat Rev Genet. 2015;16:611–22.

20. Lynch M, Field MC, Goodson HV, Malik HS, Pereira-Leal JB, Roos DS, et al. Evolutionary cell biology: two origins, one objective. Proc Natl Acad Sci U S A. 2014;111:16990–4.

21. Lyte M. Probiotics function mechanistically as delivery vehicles for neuroactive compounds: Microbial endocrinology in the design and use of probiotics. BioEssays. 2011;33:574–81.

22. Huneman P, Walsh D. Challenging the modern synthesis: Development, adaptation and inheritance. New York: Oxford University Press; 2017.

23. Pigliucci M, Müller G. Evolution: the extended synthesis. Cambridge, MA: MIT Press; 2011.

24. Wray GA, Hoekstra HE, Futuyma DJ, Lenski RE, Mackay TFC, Schluter D, et al. Does evolutionary theory need a rethink? No, all is well. Nature. 2014;514:161–4.

25. Eigen M, Schuster P. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. Die Naturwissenschaften. 1977;64:541–65.

26. Doolittle WF. Genes in pieces: Were they ever together? Nature. 1978;272:581–2.

27. Gilbert W. Why genes in pieces? Nature. 1978;271:501.

28. Irimia M, Roy SW. Origin of spliceosomal introns and alternative splicing. Cold Spring Harb Perspec Biol. 2014;6:a016071

29. de Souza SJ. Domain shuffling and the increasing complexity of biological networks. BioEssays. 2012;34:655–7.

30. Marsh JA, Teichmann SA. How do proteins gain new domains? Genome Biol. 2010;11:126.

31. Wang M, Caetano-Anolles G. The evolutionary mechanics of domain organization in proteomes and the rise of modularity in the protein world. Structure. 2009;17:66–78.

32. Koonin EV, Makarova KS. Mobile genetic elements and evolution of CRISPR-Cas systems: all the way there and back. Genome Biol Evol. 2017;9:2812–25.

33. Krupovic M, Béguin P, Koonin EV. Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. Curr Opin Microbiol. 2017;38:36–43.

34. Assis R, Kondrashov AS, Koonin EV, Kondrashov FA. Nested genes and increasing organizational complexity of metazoan genomes. Trends Genet. 2008;24:475–8.

35. Lynch M. Evolutionary diversification of the multimeric states of proteins. Proc Natl Acad Sci U S A. 2013;110:E2821–8.

36. Duboule D, Wilkins AS. The evolution of 'bricolage'. Trends Genet. 1998;14:54–9.

37. Jacob F. Evolution and tinkering. Science. 1977;196:1162.

38. Wilkins A. Between "design" and "bricolage": genetic networks, levels of selection, and adaptive evolution. Proc Natl Acad Sci U S A. 2007;104:8590–6.

39. Doolittle WF, Lukes J, Archibald JM, Keeling PJ, Gray MW. Comment on "Does constructive neutral evolution play an important role in the origin of cellular complexity?". BioEssays. 2011;33:427–9.

40. Gray MW, Lukes J, Archibald JM, Keeling PJ, Doolittle WF. Cell biology. Irremediable complexity? Science. 2010;330:920–1.

41. Lukes J, Archibald JM, Keeling PJ, Doolittle WF, Gray MW. How a neutral evolutionary ratchet can build cellular complexity. IUBMB Life. 2011;63:528–37.

42. Jain R, Rivera MC, Lake JA. Horizontal gene transfer among genomes: the complexity hypothesis. Proc Natl Acad Sci U S A. 1999;96:3801–6.

43. Lawrence JG, Roth JR. Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics. 1996;143:1843–60.

44. Promponas VJ, Ouzounis CA, Iliopoulos I. Experimental evidence validating the computational inference of functional associations from gene fusion events: a critical survey. Brief Bioinformatics. 2014;15:443–54.

45. Tsoka S, Ouzounis CA. Prediction of protein interactions: metabolic enzymes are frequently involved in gene fusion. Nat Genet. 2000;26:141–2.

46. Alon U. An introduction to systems biology: design principles of biological circuits. Florida: Chapman and Hall/CRC; 2006.

47. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science. 2002;298:824–7.

48. Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. Science. 1969;165:349–57.

49. Artzy-Randrup Y, Fleishman SJ, Ben-Tal N, Stone L. Comment on "Network motifs: simple building blocks of complex networks" and "Superfamilies of evolved and designed networks". Science. 2004;305:1107. author reply

50. Sorrells TR, Johnson AD. Making sense of transcription networks. Cell. 2015;161:714–23.

51. Carroll SB. Evolution at two levels: on genes and form. PLoS Biol. 2005;3:e245.

52. Mallarino R, Grant PR, Grant BR, Herrel A, Kuo WP, Abzhanov A. Two developmental modules establish 3D beak-shape variation in Darwin's finches. Proc Natl Acad Sci U S A. 2011;108:4057–62.

53. Peter IS, Davidson EH. Implications of developmental gene regulatory networks inside and outside developmental biology. Curr Topics Dev Biol. 2016;117:237–51.

54. Prud'homme B, Gompel N, Carroll SB. Emerging principles of regulatory evolution. Proc Natl Acad Sci U S A. 2007;104(Suppl 1):8605–12.

55. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 1999;151:1531–45.

56. Woese C. The universal ancestor. Proc Natl Acad Sci U S A. 1998;95:6854–9.

57. Coulombe-Huntington J, Xia Y. Network centrality analysis in fungi reveals complex regulation of lost and gained genes. PLoS One. 2017;12:e0169459.

58. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 2017;18:71–86.

59. Garcia-Perez JL, Widmann TJ, Adams IR. The impact of transposable elements on mammalian development. Development. 2016;143:4101–14.

60. Imbeault M, Helleboid P-Y, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature. 2017;543:550–4.

61. Sundaram V, Wang T. Transposable element mediated innovation in gene regulatory landscapes of cells: re-visiting the "gene-battery" model. BioEssays. 2018;40:1700155.

62. Ispolatov I, Yuryev A, Mazo I, Maslov S. Binding properties and evolution of homodimers in protein-protein interaction networks. Nucleic Acids Res. 2005;33:3629–35.

63. Bapteste E. The origins of microbial adaptations: how introgressive descent, egalitarian evolutionary transitions and expanded kin selection shape the network of life. Front Microbiol. 2014;5:83.

64. Bapteste E, Lopez P, Bouchard F, Baquero F, McInerney JO, Burian RM. Evolutionary analyses of non-genealogical bonds produced by introgressive descent. Proc Natl Acad Sci U S A. 2012;109:18266–72.

65. Liu J, Prindle A, Humphries J, Gabalda-Sagarra M, Asally M, Lee DY, et al. Metabolic co-dependence gives rise to collective oscillations within biofilms. Nature. 2015;523:550–4.

66. Nunes-Alves C. Biofilms: Electrifying long-range signalling. Nat Rev Microbiol. 2015;13:737.

67. Morris JJ, Lenski RE, Zinser ER. The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. MBio. 2012;3:e00036–12.

68. Sachs JL, Hollowell AC. The origins of cooperative bacterial communities. MBio. 2012;3:e00099–12.

69. Bonner JT. The origins of multicellularity. Integr Biol. 1998;1:27–36.

70. Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F, et al. Explaining microbial population genomics through phage predation. Nat Rev Microbiol. 2009;7:828–36.

71. Archibald JM. One plus one equals one: symbiosis and the evolution of complex life. Eur J Phycol. 2015;50:18.

72. Bapteste E, Anderson G. Intersecting processes are necessary explanans for evolutionary biology, but challenge retrodiction. In: Nicholson DJD, editor. Everything flows: Towards a processual philosophy of biology. Oxford: Oxford University Press. in press.

73. Koonin EV. The turbulent network dynamics of microbial evolution and the statistical Tree of Life. J Mol Evol. 2015;80:244–50.

74. Lopez-Garcia P, Zivanovic Y, Deschamps P, Moreira D. Bacterial gene import and mesophilic adaptation in archaea. Nat Rev Microbiol. 2015;13:447–56.

75. Nelson-Sathi S, Dagan T, Landan G, Janssen A, Steel M, McInerney JO, et al. Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. Proc Natl Acad Sci U S A. 2012;109:20537–42.

76. Nelson-Sathi S, Sousa FL, Roettger M, Lozada-Chavez N, Thiergart T, Janssen A, et al. Origins of major archaeal clades correspond to gene acquisitions from bacteria. Nature. 2015;517:77–80.

77. Levasseur A, Merhej V, Baptiste E, Sharma V, Pontarotti P, Raoult D. The rhizome of Lokiarchaeota illustrates the mosaicity of archaeal genomes. Genome Biol Evol. 2017;9:2635–9.

78. Akanni WA, Siu-Ting K, Creevey CJ, McInerney JO, Wilkinson M, Foster PG, et al. Horizontal gene flow from Eubacteria to Archaebacteria and what it means for our understanding of eukaryogenesis. Philos Trans R Soc Lond Ser B Biol Sci. 2015;370:20140337.

79. Martin WF, Garg S, Zimorski V. Endosymbiotic theories for eukaryote origin. Philos Trans R Soc Lond Ser B Biol Sci. 2015;370:20140330.

80. Pittis AA, Gabaldon T. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. Nature. 2016;531:101–4.

81. Archibald JM. Genomic perspectives on the birth and spread of plastids. Proc Natl Acad Sci U S A. 2015;112:10147–53.

82. Méheust R, Bhattacharya D, Pathmanathan JS, McInerney JO, Lopez P, Bapteste E. Formation of chimeric genes with essential functions at the origin of eukaryotes. BMC Biol. 2018;16:30.

83. Meheust R, Zelzion E, Bhattacharya D, Lopez P, Bapteste E. Protein networks identify novel symbiogenetic genes resulting from plastid endosymbiosis. Proc Natl Acad Sci U S A. 2016;113:3579–84.

84. Bailleul B, Berne N, Murik O, Petroutsos D, Prihoda J, Tanaka A, et al. Energetic coupling between plastids and mitochondria drives CO_2 assimilation in diatoms. Nature. 2015;524:366–9.

85. Bogumil D, Alvarez-Ponce D, Landan G, McInerney JO, Dagan T. Integration of two ancestral chaperone systems into one: the evolution of eukaryotic molecular chaperones in light of eukaryogenesis. Mol Biol Evol. 2014;31:410–8.

86. Dorrell RG, Howe CJ. Functional remodeling of RNA processing in replacement chloroplasts by pathways retained from their predecessors. Proc Natl Acad Sci U S A. 2012;109:18879–84.

87. Gavelis GS, Hayakawa S, White RA 3rd, Gojobori T, Suttle CA, Keeling PJ, et al. Eye-like ocelloids are built from different endosymbiotically acquired components. Nature. 2015;523:204–7.

88. Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, et al. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell. 2013;153:1567–78.

89. Martin W, Koonin EV. Introns and the origin of nucleus-cytosol compartmentalization. Nature. 2006;440:41–5.

90. Nowack EC, Price DC, Bhattacharya D, Singer A, Melkonian M, Grossman AR. Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of Paulinella chromatophora. Proc Natl Acad Sci U S A. 2016;113:12214–9.

91. Stairs CW, Leger MM, Roger AJ. Diversity and origins of anaerobic metabolism in mitochondria and related organelles. Philos Trans R Soc Lond Ser B Biol Sci. 2015;370:20140326.

92. Bapteste E, O'Malley MA, Beiko RG, Ereshefsky M, Gogarten JP, Franklin-Hall L, et al. Prokaryotic evolution and the tree of life are two different things. Biol Direct. 2009;4:34.

93. Koonin EV. Energetics and population genetics at the root of eukaryotic cellular and genomic complexity. Proc Natl Acad Sci U S A. 2015;112:15777–8.

94. Martin W, Muller M. The hydrogen hypothesis for the first eukaryote. Nature. 1998;392:37–41.

95. Williams TA, Embley TM. Changing ideas about eukaryotic origins. Philos Trans R Soc Lond Ser B Biol Sci. 2015;370:20140318.

96. Simplicity BA. The Stanford Encyclopedia of Philosophy: Metaphysics Research Lab, Stanford University; 2016. https://plato.stanford.edu/archives/win2016/entries/simplicity/

97. Allen JF. Why chloroplasts and mitochondria retain their own genomes and genetic systems: Colocation for redox regulation of gene expression. Proc Natl Acad Sci U S A. 2015;112:10231–8.

98. Godfrey-Smith P. Reproduction, symbiosis, and the eukaryotic cell. Proc Natl Acad Sci U S A. 2015;112:10120–5.

99. Karnkowska A, Vacek V, Zubacova Z, Treitli SC, Petrzelkova R, Eme L, et al. A eukaryote without a mitochondrial organelle. Curr Biol. 2016;26:1274–84.

100. Gilbert SF, Sapp J, Tauber AI. A symbiotic view of life: we have never been individuals. Q Rev Biol. 2012;87:325–41.

101. Moran NA, Sloan DB. The hologenome concept: helpful or hollow? PLoS Biol. 2015;13:e1002311.

102. Bouchard F. Understanding colonial traits using symbiosis research and ecosystem ecology. Biol Theory. 2009;4:240–6.

103. Selosse MA, Bessis A, Pozo MJ. Microbial priming of plant and animal immunity: symbionts as developmental signals. Trends Microbiol. 2014;22:607–13.

104. Theis KR, Dheilly NM, Klassen JL, Brucker RM, Baines JF, Bosch TC, et al. Getting the hologenome concept right: an eco-evolutionary framework for hosts and their microbiomes. mSystems. 2016;1:e00028–16.

105. Theis KR, Venkataraman A, Dycus JA, Koonter KD, Schmitt-Matzen EN, Wagner AP, et al. Symbiotic bacteria appear to mediate hyena social odors. Proc Natl Acad Sci U S A. 2013;110:19832–7.

106. Sharon G, Segal D, Ringo JM, Hefetz A, Zilber-Rosenberg I, Rosenberg E. Commensal bacteria play a role in mating preference of Drosophila melanogaster. Proc Natl Acad Sci U S A. 2010;107:20051–6.

107. Bosch TC, McFall-Ngai MJ. Metaorganisms as the new frontier. Zoology. 2011;114:185–90.

108. Bordenstein SR, Theis KR. Host biology in light of the microbiome: ten principles of holobionts and hologenomes. PLoS Biol. 2015;13:e1002226.

109. Bosch TC. Rethinking the role of immunity: lessons from Hydra. Trends Immunol. 2014;35:495–502.

110. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJ, Relman DA. The application of ecological theory toward an understanding of the human microbiome. Science. 2012;336:1255–62.

111. Barr JJ, Auro R, Furlan M, Whiteson KL, Erb ML, Pogliano J, et al. Bacteriophage adhering to mucus provide a non-host-derived immunity. Proc Natl Acad Sci U S A. 2013;110:10771–6.

112. Rosenberg E, Sharon G, Zilber-Rosenberg I. The hologenome theory of evolution contains Lamarckian aspects within a Darwinian framework. Environmental Microbiol. 2009;11:2959–62.

113. Ley RE. The gene-microbe link. Nature. 2015;518:S7.

114. Tsuchida T, Koga R, Horikawa M, Tsunoda T, Maoka T, Matsumoto S, et al. Symbiotic bacterium modifies aphid body color. Science. 2010;330:1102–4.

115. Bravo JA, Forsythe P, Chew MV, Escaravage E, Savignac HM, Dinan TG, et al. Ingestion of Lactobacillus strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. Proc Natl Acad Sci U S A. 2011;108:16050–5.

116. Dupressoir A, Lavialle C, Heidmann T. From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. Placenta. 2012;33:663–71.

117. Emera D, Casola C, Lynch VJ, Wildman DE, Agnew D, Wagner GP. Convergent evolution of endometrial prolactin expression in primates, mice, and elephants through the independent recruitment of transposable elements. Mol Biol Evol. 2012;29:239–47.

118. Ezenwa VO, Gerardo NM, Inouye DW, Medina M, Xavier JB. Microbiology. Animal behavior and the microbiome. Science. 2012;338:198–9.

119. Scarborough CL, Ferrari J, Godfray HC. Aphid protected from pathogen by endosymbiont. Science. 2005;310:1781.

120. Hanski I. Metapopulation ecology. Oxford: Oxford University Press; 1999.

121. Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, et al. The metacommunity concept: a framework for multi-scale community ecology. Ecol Lett. 2004;7:601–13.

122. Ricklefs RE. Disintegration of the ecological community. Am Nat. 2008;172:741–50.

123. Darwin CA. On the origin of species by means of natural selection. London: John Murray; 1859.

124. Caporael L, Griesemer J, Wimsatt W. Scaffolding in evolution, culture, and cognition. Cambridge: MIT Press; 2013.

125. Laland K, Matthews B, Feldman MW. An introduction to niche construction theory. Evol Ecol. 2016;30:191–202.

126. Rogers RL, Hartl DL. Chimeric genes as a source of rapid evolution in Drosophila melanogaster. Mol Biol Evol. 2012;29:517–29.

127. Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc Lond B Biol Sci. 1979;205:581–98.

128. Sole RV, Valverde S. Are network motifs the spandrels of cellular complexity? Trends Ecol Evol. 2006;21:419–22.

129. O'Malley MA, Soyer OS, Siegal ML. A philosophical perspective on evolutionary systems biology. Biol Theory. 2015;10:6–17.

130. Soyer OS, O'Malley MA. Evolutionary systems biology: what it is and why it matters. BioEssays. 2013;35:696–705.

131. Karimpour-Fard A, Hunter L, Gill RT. Investigation of factors affecting prediction of protein-protein interaction networks by phylogenetic profiling. BMC Genomics. 2007;8:393.

132. Koch C, Konieczka J, Delorey T, Lyons A, Socha A, Davis K, et al. Inference and Evolutionary Analysis of Genome-Scale Regulatory Networks in Large Phylogenies. Cell Syst. 2017;4:543–58. e8

133. Shahdoust M, Pezeshk H, Mahjub H, Sadeghi M. F-MAP: a Bayesian approach to infer the gene regulatory network using external hints. PLoS One. 2017;12:e0184795.

134. Simonsen M, Maetschke SR, Ragan MA. Automatic selection of reference taxa for protein-protein interaction prediction with phylogenetic profiling. Bioinformatics. 2012;28:851–7.

135. Spanier KI, Jansen M, Decaestecker E, Hulselmans G, Becker D, Colbourne JK, et al. Conserved transcription factors steer growth-related genomic programs in Daphnia. Genome Biol Evol. 2017;9:1821–42.

136. Wang P, Yu X, Lü J. Identification and evolution of structurally dominant nodes in protein-protein interaction networks. IEEE Trans Biomed Circuits Syst. 2014;8:87–97.

137. Ruprecht C, Vaid N, Proost S, Persson S, Mutwil M. Beyond genomics: studying evolution with gene coexpression networks. Trends Plant Sci. 2017; 22:298–307.

138. Netotea S, Sundell D, Street NR, Hvidsten TR. ComPlEx: conservation and divergence of co-expression networks in A. thaliana, Populus and O. sativa. BMC Genomics. 2014;15:106.

139. Conant GC, Wolfe KH. Functional partitioning of yeast co-expression networks after genome duplication. PLoS Biol. 2006;4:e109.

140. You Q, Xu W, Zhang K, Zhang L, Yi X, Yao D, et al. ccNET: Database of co-expression networks with functional modules for diploid and polyploid Gossypium. Nucleic Acids Res. 2017;45:D1090–D9.

141. Tang J, Lin J, Li H, Li X, Yang Q, Cheng Z-M, et al. Characterization of CIPK family in Asian Pear (Pyrus bretschneideri Rehd) and co-expression analysis related to salt and osmotic stress responses. Front Plant Sci. 2016;7:1361.

142. Siwo GH, Tan A, Button-Simons KA, Samarakoon U, Checkley LA, Pinapati RS, et al. Predicting functional and regulatory divergence of a drug resistance transporter gene in the human malaria parasite. BMC Genomics. 2015;16:115.

143. Hu G, Hovav R, Grover CE, Faigenboim-Doron A, Kadmon N, Page JT, et al. Evolutionary conservation and divergence of gene coexpression networks in Gossypium (cotton) seeds. Genome Biol Evol. 2016;8:3765–83.

144. Lu X, Li Q-T, Xiong Q, Li W, Bi Y-D, Lai Y-C, et al. The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. Plant J. 2016;86:530–44.

145. Zandveld J, van den Heuvel J, Mulder M, Brakefield PM, Kirkwood TBL, Shanley DP, et al. Pervasive gene expression responses to a fluctuating diet in Drosophila melanogaster: The importance of measuring multiple traits to decouple potential mediators of life span and reproduction. Evolution. 2017;71:2572–83.

146. Halfon MS. Perspectives on gene regulatory network evolution. Trends Genet. 2017;33:436–47.

147. Simakov O, Kawashima T. Independent evolution of genomic characters during major metazoan transitions. Dev Biol. 2017;427:179–92.

148. Wang P, Zhao D, Rockowitz S, Zheng D. Divergence and rewiring of regulatory networks for neural development between human and other species. Neurogenesis (Austin). 2016;3:e1231495.

149. Masalia RR, Bewick AJ, Burke JM. Connectivity in gene coexpression networks negatively correlates with rates of molecular evolution in flowering plants. PLoS One. 2017;12:e0182289.

150. Muñoz A, Santos Muñoz D, Zimin A, Yorke JA. Evolution of transcriptional networks in yeast: alternative teams of transcriptional factors for different species. BMC Genomics. 2016;17:826.

151. Kacharia FR, Millar JA, Raghavan R. Emergence of New sRNAs in Enteric Bacteria is Associated with Low Expression and Rapid Evolution. J Mol Evol. 2017;84:204–13.

152. Kim HS, Mittenthal JE, Caetano-Anollés G. MANET: tracing evolution of protein architecture in metabolic networks. BMC Bioinformatics. 2006;7:351.

153. Leyn SA, Suvorova IA, Kazakov AE, Ravcheev DA, Stepanova VV, Novichkov PS, et al. Comparative genomics and evolution of transcriptional regulons inProteobacteria. Microb Genom. 2016;2:e000061.

154. Liang C, Luo J, Song D. Network simulation reveals significant contribution of network motifs to the age-dependency of yeast protein-protein interaction networks. Mol BioSyst. 2014;10:2277–88.

155. Mustafin ZS, Lashin SA, Matushkin YG, Gunbin KV, Afonnikov DA. Orthoscape: a cytoscape application for grouping and visualization KEGG based gene networks by taxonomy and homology principles. BMC Bioinformatics. 2017;18:1427.

156. Thompson JR, Erkenbrack EM, Hinman VF, McCauley BS, Petsios E, Bottjer DJ. Paleogenomics of echinoids reveals an ancient origin for the double-negative specification of micromeres in sea urchins. Proc Natl Acad Sci U S A. 2017;114:5870–7.

157. Corel E, Lopez P, Méheust R, Bapteste E. Network-thinking: graphs to analyze microbial complexity and evolution. Trends Microbiol. 2016;24: 224–37.

158. Friedlander T, Prizak R, Barton NH, Tkačik G. Evolution of new regulatory functions on biophysically realistic fitness landscapes. Nat Commun. 2017;8:216.

159. Gouy A, Daub JT, Excoffier L. Detecting gene subnetworks under selection in biological pathways. Nucleic Acids Res. 2017;45:e149.

160. MacKintosh C, DEK F. Recent advances in understanding the roles of whole genome duplications in evolution. F1000Res. 2017;6:1623.

161. Nguyen Ba AN, Strome B, Osman S, Legere E-A, Zarin T, Moses AM. Parallel reorganization of protein function in the spindle checkpoint pathway through evolutionary paths in the fitness landscape that appear neutral in laboratory experiments. PLoS Genet. 2017;13:e1006735.

162. Alvarez-Ponce D, Feyertag F, Chakraborty S. Position matters: network centrality considerably impacts rates of protein evolution in the human protein-protein interaction network. Genome Biol Evol. 2017;9:1742–56.

163. Cohen O, Gophna U, Pupko T. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. Mol Biol Evol. 2011;28:1481–9.

164. Raymond J, Segrè D. The effect of oxygen on biochemical networks and the evolution of complex life. Science. 2006;311:1764–7.

165. Wolf YI, Carmel L, Koonin EV. Unifying measures of gene function and evolution. Proc Biol Sci. 2006;273:1507–15.

166. Hase T, Niimura Y, Tanaka H. Difference in gene duplicability may explain the difference in overall structure of protein-protein interaction networks among eukaryotes. BMC Evol Biol. 2010;10:358.

167. Xu K, Bezakova I, Bunimovich L, Yi SV. Path lengths in protein-protein interaction networks and biological complexity. Proteomics. 2011;11:1857–67.

168. Prachumwat A, Li W-H. Protein function, connectivity, and duplicability in yeast. Mol Biol Evol. 2006;23:30–9.

169. Peterson GJ, Pressé S, Peterson KS, Dill KA. Simulated evolution of protein-protein interaction networks with realistic topology. PLoS One. 2012;7: e39052.

170. Pawlowski PH, Kaczanowski S, Zielenkiewicz P. A kinetic model of the evolution of a protein interaction network. BMC Genomics. 2013;14:172.

171. Ruprecht C, Proost S, Hernandez-Coronado M, Ortiz-Ramirez C, Lang D, Rensing SA, et al. Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules. Plant J. 2017;90:447–65.

172. Zhao Y, Mooney SD. Functional organization and its implication in evolution of the human protein-protein interaction network. BMC Genomics. 2012;13:150.

173. Akinola RO, Mazandu GK, Mulder NJ. A quantitative approach to analyzing genome reductive evolution using protein-protein interaction networks: a case study of Mycobacterium leprae. Front Genet. 2016;7:39.

174. Briones-Moreno A, Hernández-García J, Vargas-Chávez C, Romero-Campero FJ, Romero JM, Valverde F, et al. Evolutionary analysis of DELLA-associated transcriptional networks. Front Plant Sci. 2017;8:626.

175. Hahn MW, Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol. 2005;22:803–6.

176. Hansen BO, Vaid N, Musialak-Lange M, Janowski M, Mutwil M. Elucidating gene function and function evolution through comparison of co-expression networks of plants. Front Plant Sci. 2014;5:394.

177. Kelliher CM, Leman AR, Sierra CS, Haase SB. Investigating conservation of the cell-cycle-regulated transcriptional program in the fungal pathogen, Cryptococcus neoformans. PLoS Genet. 2016;12:e1006453.

178. Martinez-Pastor M, Tonner PD, Darnell CL, Schmid AK. Transcriptional regulation in Archaea: from individual genes to global regulatory networks. Annu Rev Genet. 2017;51:143–70.

179. Phan HTT, Sternberg MJE. PINALOG: a novel approach to align protein interaction networks–implications for complex detection and function prediction. Bioinformatics. 2012;28:1239–45.

180. Romero-Campero FJ, Perez-Hurtado I, Lucas-Reina E, Romero JM, Valverde F. ChlamyNET: a Chlamydomonas gene co-expression network reveals global properties of the transcriptome and the early setup of key co-expression patterns in the green lineage. BMC Genomics. 2016;17:227.

181. Tamames J, Moya A, Valencia A. Modular organization in the reductive evolution of protein-protein interaction networks. Genome Biol. 2007;8:R94.

182. Wang D, He F, Maslov S, Gerstein M. DREISS: using state-space models to infer the dynamics of gene expression driven by external and internal regulatory networks. PLoS Comput Biol. 2016;12:e1005146.

183. Aguilera F, McDougall C, Degnan BM. Co-option and de novo gene evolution underlie molluscan shell diversity. Mol Biol Evol. 2017;34:779–92.

184. Auman T, Chipman AD. The evolution of gene regulatory networks that define arthropod body plans. Integr Comp Biol. 2017;57:523–32.

185. Mateos JL, Tilmes V, Madrigal P, Severing E, Richter R, Rijkenberg CWM, et al. Divergence of regulatory networks governed by the orthologous transcription factors FLC and PEP1 in Brassicaceae species. Proc Natl Acad Sci U S A. 2017;114:E11037–E46.

186. Renvoisé E, Kavanagh KD, Lazzari V, Häkkinen TJ, Rice R, Pantalacci S, et al. Mechanical constraint from growing jaw facilitates mammalian dental diversity. Proc Natl Acad Sci U S A. 2017;114:9403–8.

187. Cary GA, Cheatle Jarvela AM, Francolini RD, Hinman VF. Genome-wide use of high- and low-affinity Tbrain transcription factor binding sites during echinoderm development. Proc Natl Acad Sci U S A. 2017;114:5854–61.

188. Noble R, Noble D. Was the watchmaker blind? Or was she one-eyed? Biology (Basel). 2017;6:47.

189. Capela D, Marchetti M, Clérissi C, Perrier A, Guetta D, Gris C, et al. Recruitment of a lineage-specific virulence regulatory pathway promotes intracellular infection by a plant pathogen experimentally evolved into a legume symbiont. Mol Biol Evol. 2017;34:2503–21.

190. Orsini L, Brown JB, Shams Solari O, Li D, He S, Podicheti R, et al. Early transcriptional response pathways in Daphnia magna are coordinated in networks of crustacean-specific genes. Mol Ecol. 2018;27:886–97.

191. Dohrmann J, Puchin J, Singh R. Global multiple protein-protein interaction network alignment by combining pairwise network alignments. BMC Bioinformatics. 2015;16(Suppl 13):S11.

192. Emmert-Streib F, Dehmer M, Shi Y. Fifty years of graph matching, network alignment and network comparison. Inf Sci. 2016;346:180–97.

193. Kelley BP, Yuan B, Lewitter F, Sharan R, Stockwell BR, Ideker T. PathBLAST: a tool for alignment of protein interaction networks. Nucleic Acids Res. 2004; 32:W83–8.

194. Ni B, Ghosh B, Paldy FS, Colin R, Heimerl T, Sourjik V. Evolutionary remodeling of bacterial motility checkpoint control. Cell Rep. 2017;18:866–77.

195. Huisman M. Imputation of missing network data: some simple procedures. In: Alhajj R, Rokne J, editors. Encyclopedia of Social Network Analysis and Mining. New York: Springer New York; 2014. p. 707–15.

196. Ogundijo OE, Elmas A, Wang X. Reverse engineering gene regulatory networks from measurement with missing values. EURASIP J Bioinform Syst Biol. 2016;2017:2.

197. Shao M, Zhou S, Guan J. Revisiting topological properties and models of protein-protein interaction networks from the perspective of dataset evolution. IET Syst Biol. 2015;9:113–9.

198. Doolittle WF, Inkpen SA. Processes and patterns of interaction as units of selection: An introduction to ITSNTS thinking. Proc Natl Acad Sci U S A. 2018;115:4006–14.

199. Wong DCJ, Matus JT. Constructing integrated networks for identifying new secondary metabolic pathway regulators in grapevine: recent applications and future opportunities. Front Plant Sci. 2017;8:505.

200. Corel E, Méheust R, Watson AK, McInerney JO, Lopez P, Bapteste E. Bipartite network analysis of gene sharings in the microbial world. Mol Biol Evol. 2017;35:899–913.

201. Seligmann H, Raoult D. Unifying view of stem–loop hairpin RNA as origin of current and ancient parasitic and non-parasitic RNAs, including in giant viruses. Curr Opin Microbiol. 2016;31:1–8.

202. Stoltzfus A. Constructive neutral evolution: exploring evolutionary theory's curious disconnect. Biol Direct. 2012;7:35.

203. Gillespie JH. Population genetics: A concise guide. Baltimore: The Johns Hopkins University Press; 2004.

204. Barker G, Desjardins E, Pearce T. Entangled Life: Organism and Environment in the Biological and Social Sciences. Dordrecht: Springer; 2013.

205. Bonduriansky R. Rethinking heredity, again. Trends Ecol Evol. 2012;27:330–6.

206. Lehmann L. The adaptive dynamics of niche constructing traits in spatially subdivided populations: evolving posthumous extended phenotypes. Evolution. 2008;62:549–66.

207. Bruno JF, Stachowicz JJ, Bertness MD. Inclusion of facilitation into ecological theory. Trends Ecol Evol. 2003;18:119–25.

208. Beatty J. Replaying life's tape. J Philosophy. 2006;103:336–62.

209. Turner D. Historical contingency and the explanation of evolutionary trends. In: Malaterre C, Braillard PA, editors. Biological explanation: An enquiry into the diversity of explanatory patterns in the life sciences. Dordrecht: Springer; 2015. p. 73–90.

210. O'Hara RJ. Population thinking and tree thinking in systematics. Zool Scr. 1997;26:323–9.

211. Woese CR. A new biology for a new century. Microbiol Mol Biol Rev. 2004; 68:173–86.

212. Lange M. Really statistical explanations and genetic drift. Philosophy Sci. 2013;80:169–88.

213. Huneman P. Diversifying the picture of explanations in biological sciences: Ways of combining topology with mechanisms. Synthese. 2018;195:115–46.

214. Jones N. Bowtie structures, pathway diagrams, and topological explanation. Erkenntnis. 2014;89:1355–555.

215. Lane N, Martin WF, Raven JA, Allen JF. Energy, genes and evolution: introduction to an evolutionary synthesis. Philos Trans R Soc Lond Ser B Biol Sci. 2013;368:20120253.

216. Sousa FL, Thiergart T, Landan G, Nelson-Sathi S, Pereira IA, Allen JF, et al. Early bioenergetic evolution. Philos Trans R Soc Lond Ser B Biol Sci. 2013; 368:20130088.

217. Nagel E. The structure of science. Harcourt, Brace & World: New-York; 1961.

218. Booth A, Mariscal C, Doolittle WF. The Modern Synthesis in the light of microbial genomics. Annu Rev Microbiol. 2016;70:279–97.

219. Gilbert SF, Opitz JM, Raff RA. Resynthesizing evolutionary and developmental biology. Dev Biol. 1996;173:357–72.

220. Kokko H, Chaturvedi A, Croll D, Fischer MC, Guillaume F, Karrenberg S, et al. Can evolution supply what ecology demands? Trends Ecol Evol. 2017;32: 187–97.

221. Cancherini DV, França GS, de Souza SJ. The role of exon shuffling in shaping protein-protein interaction networks. BMC Genomics. 2010;11(Suppl 5):S11.

222. Amin SA, Hmelo LR, van Tol HM, Durham BP, Carlson LT, Heal KR, et al. Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. Nature. 2015;522:98–101.

223. Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, et al. Communication between viruses guides lysis-lysogeny decisions. Nature. 2017;541:488–93.

224. Dorrell RG, Gile G, McCallum G, Meheust R, Bapteste EP, Klinger CM, et al. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. elife. 2017;6:e23717.

scientific reports

OPEN

The role of borosilicate glass in Miller–Urey experiment

Joaquín Criado-Reyes[1], Bruno M. Bizzarri[2], Juan Manuel García-Ruiz[1 ✉],
Raffaele Saladino[2 ✉] & Ernesto Di Mauro[2]

We have designed a set of experiments to test the role of borosilicate reactor on the yielding of the Miller–Urey type of experiment. Two experiments were performed in borosilicate flasks, two in a Teflon flask and the third couple in a Teflon flask with pieces of borosilicate submerged in the water. The experiments were performed in CH_4, N_2, and NH_3 atmosphere either buffered at pH 8.7 with NH_4Cl or unbuffered solutions at pH ca. 11, at room temperature. The Gas Chromatography-Mass Spectroscopy results show important differences in the yields, the number of products, and molecular weight. In particular, a dipeptide, multi-carbon dicarboxylic acids, PAHs, and a complete panel of biological nucleobases form more efficiently or exclusively in the borosilicate vessel. Our results offer a better explanation of the famous Miller's experiment showing the efficiency of borosilicate in a triphasic system including water and the reduced Miller–Urey atmosphere.

The 1953's publication of the Miller–Urey experiment opened the door to the scientific investigation of the origin of life[1]. In this brilliant experiment, Miller and Urey demonstrated that electrical sparking a mixture of methane, ammonia, and hydrogen in the presence of water produces amino acids within a variety of organic compounds. The impact of these results was so high that its mind-opening relevance hardly fades over time[2]. Different gas mixtures have been explored[3–7], and the yielding and molecular diversity were confirmed with modern analytical techniques[8], including original sample remnants of early Miller experiments[9,10]. Variations of the original Miller apparatus have been used, but the experiments were always performed within borosilicate flasks. Interestingly, the initial pH of most of the canonical mixtures aiming to mimic the early Earth atmosphere in Miller–Urey experiments is highly alkaline. As reported by Miller[1,3], under these alkaline conditions, silica dissolves: the higher the pH and temperature, the higher the solubility of silica (Fig. S1). Therefore, it could be expected that upon contact of the alkaline water with the inner wall of the borosilicate flask, even this reinforced glass will slightly dissolve releasing silica and traces of other metal oxides, offering silanol groups to the gas phase and the liquid water and vapor. Motivated by the biomimetic role of silica in mineral self-organized structures, such as silica-carbonate biomorphs[11–13] and its catalytic role in prebiotic chemistry[14,15], we designed a set of experiments to test the possible influence of silica on the classical Miller experiments.

Results

Figure 1 shows the experimental concept. Three types of experiments were carried out under two different chemical conditions, one unbuffered with a starting pH value of ca. 11, the other buffered at pH 8.7. One of the experiments was performed in a borosilicate reactor (hereafter BSR unbuffered and BSRB buffered) as used in Miller-type experiments. A second was performed in a Teflon reactor (TFR unbuffered and TFRB buffered), a third in a Teflon reactor with centimeter pieces of borosilicate glass submerged in the water (TFBSR unbuffered and TFBSR/B buffered). After proceeding with the electrical discharges, the differences in color of the collected samples were visually evident (Fig. S2). In what follows, we describe the results of these experiments.

We used a single flask Miller-Urey apparatus where electrodes, water, and the components of the atmosphere were joined in one single reaction flask made either of borosilicate or Teflon. The borosilicate flask (Duran) had a volume of 3 L, the Teflon flask of 1.5 L (Fig. S3). A Tesla coil provided the 30 kV to ignite the electric arc between the tungsten electrodes. The flasks were filled with water to a volume of 200 mL, so the sparking took place in the gas phase (Figure S4 and S5). All the experiments were performed at room temperature, with a water vapor pressure of ca. 24 mbars, to remove thermal effects for a more effective comparison (see further details in SI# 1). We selected one of the most effective Miller atmospheres made of ammonia (200 ± 20 mbar), methane (200 ± 20 mbar), and nitrogen (100 ± 20 mbar). Ammonia and nitrogen are considered ubiquitous components

[1]Laboratorio de Estudios Cristalográficos, Instituto Andaluz de Ciencias de la Tierra, Consejo Superior de Investigaciones Científicas, Universidad de Granada, Avenida de las Palmeras 4, Armilla, 18100 Granada, Spain. [2]Ecological and Biological Sciences Department (DEB), University of Tuscia, Via S. Camillo de Lellis snc, 01100 Viterbo, Italy. ✉email: juanmanuel.garcia@csic.es; saladino@unitus.it
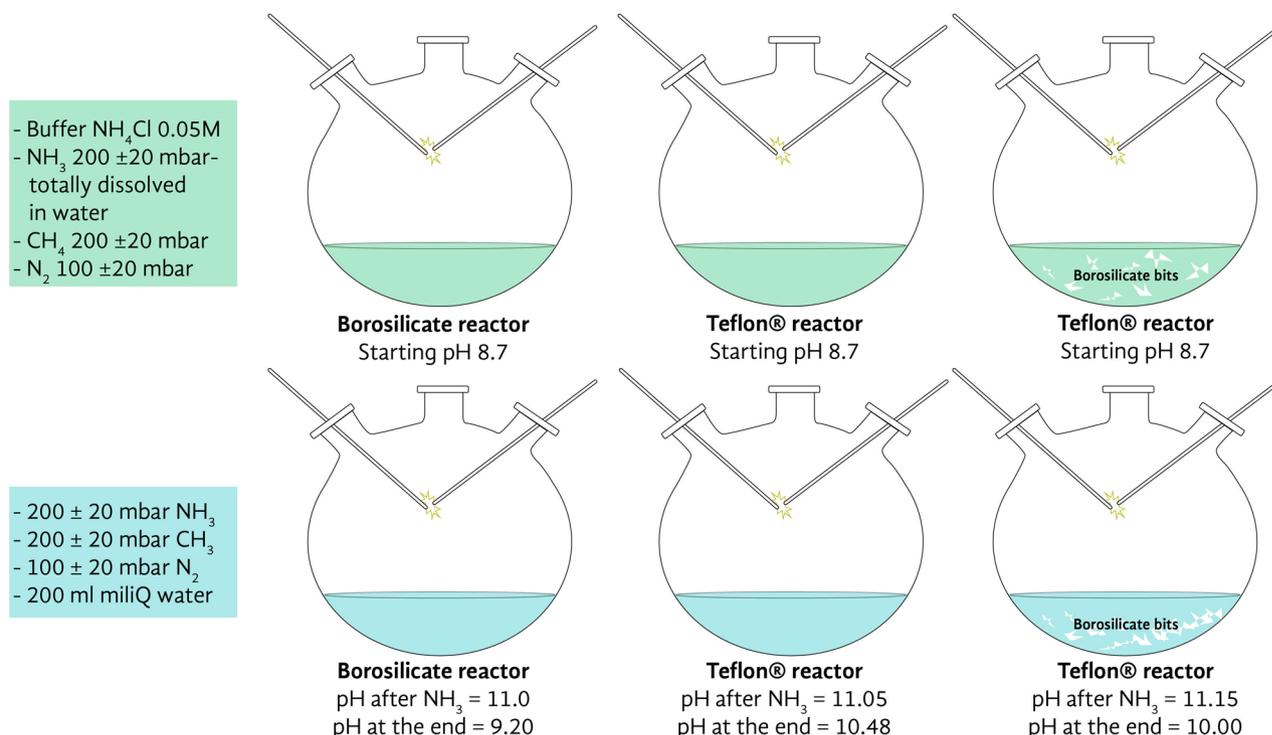
Figure 1. The experimental design. Six electric discharge experiments were performed in two different flasks, one made of borosilicate glass, the other of Teflon. Three experiments were performed to test the effect of the borosilicate glass. One in the borosilicate reactor, one in the Teflon reactor, a third one in the Teflon reactor containing pieces of borosilicate glass. The three experiments were repeated with NH₄Cl buffer at pH 8.7 and without buffer at pH ca. 11.

of the early atmosphere[16,17]. The initial pH value of the water was ca. 11.1, and it decreased during the run by almost two units in the borosilicate flask and one unit in the Teflon reactor. The experiments with the NH₄Cl buffer were aimed to keep the pH constant in the region where the speciation is not only $H_4SiO_4^=$ silicic acid but also has a small contribution of $H_3SiO_4^-$. They also mimic the presumptive presence of the ammonium ion NH_4^+ in the primitive ocean[18] and optimize the synthesis of amino acids by the Strecker condensation[19]. The crude was analyzed by gas chromatography associated to mass spectrometry (GC–MS) after derivatization of the sample to corresponding trimethylsilyl ethers (TMS), the yield of reaction products was reported as both micrograms of product per 1.0 mg of crude and mg of product per total amount of crude (SI#1). The structure of reaction products was tentatively assigned by comparison of the mass fragmentation spectra with the original one deposited in the database and further confirmed, when the similarity index was lower than 98%, by the co-injection method with original standards (SI#3). The most abundant reaction products are described in Fig. 2 and Table S2 (buffered condition) and Table S3 (unbuffered condition), the mass to charge (*m/z*) ratio values and relative peak abundances of products are in SI#2 (Table S4), while GC chromatograms and original *m/z* fragmentation spectra are in SI#3 and SI#4, respectively. As shown in Fig. 2 and Tables S2-S3, a large panel of elemental prebiotic chemical precursors (ECP) **1–4**, amino acids and alkyl amines **5–24**, carboxylic acids **25–35**, RNA and DNA nucleobases **36–40**, and aromatic and heteroaromatic derivatives **41–48** were tentatively assigned in different yield and selectivity depending on the specific experimental conditions. The total yield of compounds **1–48** grouped per chemical class similarity is reported in Table 1. In spite of these circumstantial indications, in our opinion the possibility exists that there could still be in principle an effect exerted by the size and shape of the reactor, even though we consider unlikely that these could significantly modify the selectivity and efficiency of the observed reaction pathways. The same applies for the electrode gap variation. The correlation between product distribution and variation of electrode geometry has been discussed[7].

Overall, these results confirm the visual assessment that the brown broth obtained in the borosilicate experiments contained much more organic compounds than those of the Teflon experiments, irrespective of the buffering (Fig. S2; Table 1, entry 7). A larger panel of reaction products was obtained in borosilicate with respect to Teflon alone (48 compounds versus 31; Tables S2-S3), and several amino acids, a dipeptide, carboxylic acids and aromatic miscellanea (for a total of 17 compounds) were produced only in the presence of borosilicate (Tables S2-S3) (Fig. 3A).

Borosilicate increased the yield of ECP **1–4** relative to Teflon alone (Table 1, entry 1). Ab initio atomistic simulation of the Miller-Urey experiment postulated the barrier-less formation of **1** and **2** from a reducing atmosphere[20], and traces of these compounds of key importance in prebiotic chemistry[21] were recently detected by mimicking a meteoritic impact in the pristine atmosphere[22]. DAMN **4** is a common intermediate in the synthesis of nucleobases from HCN and **1**, while **3** is a component of the organic pool in the primitive Earth[23]. A total of 17 amino acids **5–21**, a dipeptide **22**, and two simple amines **23–24**, were detected in the crude (Fig. 2).
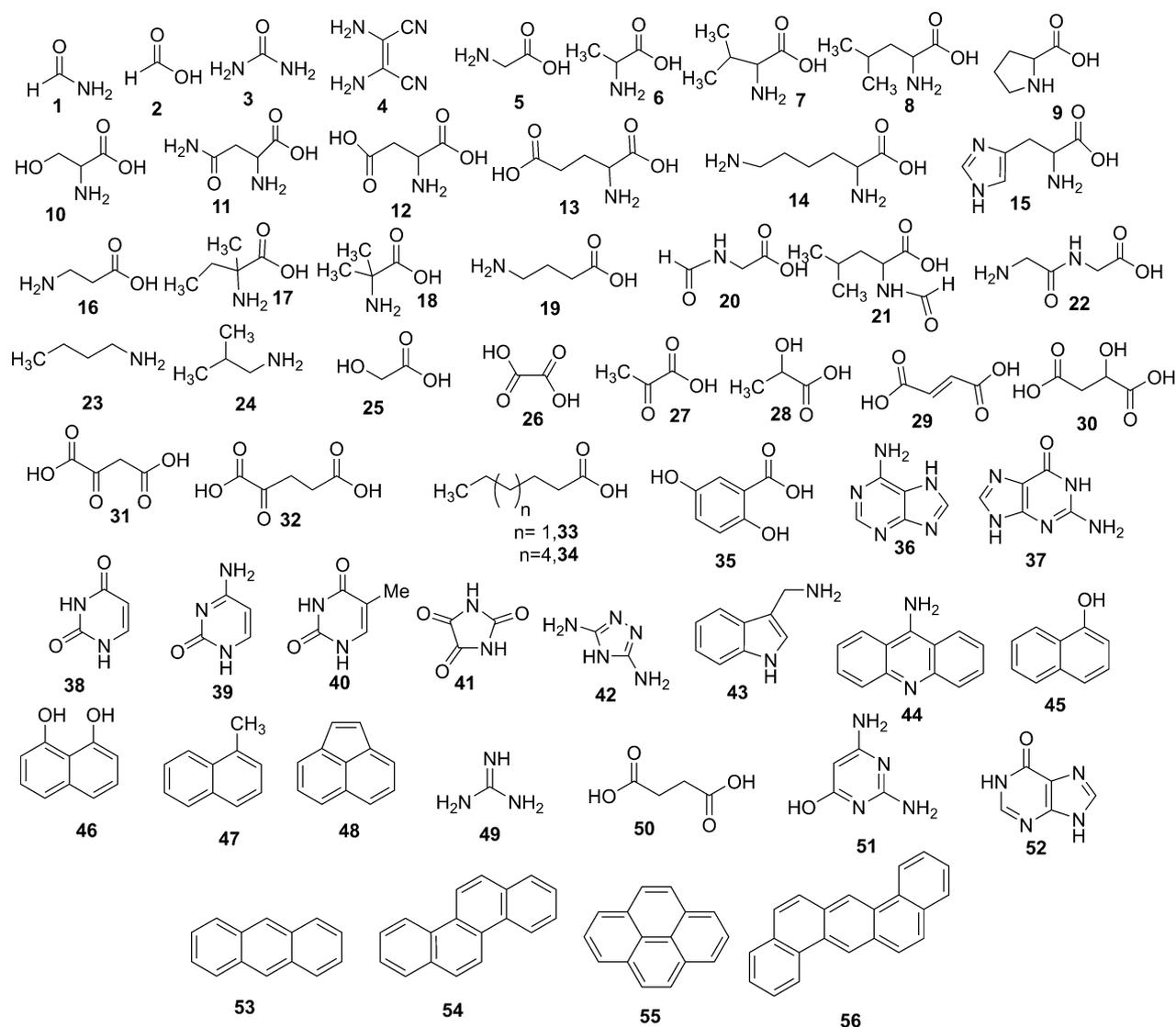
Figure 2. Overall view of the organic compounds produced during the six electric discharge experiments performed in borosilicate glass and Teflon flasks. Formamide **1**, formic acid **2**, urea **3**, diaminomaleonitrile **4**, glycine **5**, alanine **6**, valine **7**, leucine **8**, proline **9**, serine **10**, asparagine **11**, aspartic acid **12**, glutamic acid **13**, lysine **14**, histidine **15**, β-alanine **16**, iso-valine **17**, α-amino isobutyric **18**, γ-aminoisobutyric acid **19**, N-formyl glycine **20**, N-formyl leucine **21**, glycylglycine **22**, n-butanamine **23**, 2-methylpropanamine **24**, glycolic acid **25**, oxalic acid **26**, pyruvic acid **27**, lactic acid **28**, fumaric acid **29**, malic acid **30**, oxaloacetic acid **31**, α-ketoglutaric acid **32**, n-hexanoic acid **33**, n-nonanoic acid **34**, gentisic acid **35**, adenine **36**, guanine **37**, uracil **38**, cytosine **39**, thymine **40**, parabanic acid **41**, 3,5-diamino-1,2,4-triazole **42**, 1(H)-indole-3-methanamine **43**, 9-acridinamine **44**, 1-hydroxynaphtalene (naphtol) **45**, 1,8-dihydroxynaphtalene **46**, methylnaphthalene **47**, acenaphthylene **48**, guanidine **49**, succinic acid **50**, 2,4-diamino-6-hydroxypyrimidine **51**, hypoxanthine **52**, anthracene **53**, crysene **54**, pyrene **55**, and dibenz(a,h)anthracene **56**.

The total yield of amino acids was higher in the borosilicate flasks than in Teflon alone (Table 1, entry 2). In addition, amino acids **6–7**, **12–13**, and **17**, and the dipeptide **22**, formed exclusively in the presence of borosilicate (Fig. 2, Tables S2-S3).

The synthesis of **22**, as well as that of formylated amino acids **20–21** (Fig. 2), is of prebiotic relevance and was probably favored by the formation of carbodiimide from **1**, a borosilicate-catalyzed process[24]. Once formed, carbodiimide can activate amino acids towards the formation of the peptide bond with contemporaneous release of urea[25]. Carboxylic acids **25–35** (from C-1 to C-9) were also tentatively identified in the reaction mixture (Fig. 2), the highest total yield being obtained in the presence of borosilicate (Table S3, entry 3). Carboxylic acids **25**, **30–31**, **32**, and **35** were absent in the experiment performed in Teflon alone (Tables S1-S2). The beneficial role of borosilicate was further confirmed in the synthesis of nucleobases. In this latter case, borosilicate systems afforded the complete set of nucleobases **36–40**, while only **36**, **39**, and **40** were detected in the Teflon flask (Tables S2-S3). Again, the total yield of nucleobases was highest in the presence of borosilicate (Table 1,

Entry	Class	BSRB	TFRB	TFBSR/B*	BSR	TFR	TFBSR*
		Yield (μg product/1.0 mg of crude)					
1	ECP	122.96	31.29	100.91	129.64	35.46	90.18
2	Amino acids	159.45	27.1	60.53	111.19	51.20	87.41
3	Carboxylic acids	36.65	11.91	26.49	46.3	28.21	80.89
4	Nucleobases	14.01	7.34	5.83	16.3	4.69	14.02
5	Aromatic miscellanea	26.95	7.14	10.86	23.07	48.25	33.58
6	Amines	33.80	33.50	34.79	77.07	78.19	69.06
7	Total amount	393.82	118.28	239.41	403.57	246	375.14

Table 1. Total yield of products grouped for chemical class: ECP elemental prebiotic chemical precursors compounds, amino acids, carboxylic acids, nucleobases, aromatic miscellanea, amines. *BSRB* borosilicate in buffer, *TFRB* Teflon in buffer, *TFBSR/B* Teflon in buffer in the presence of pieces of borosilicate, *BSR* borosilicate without buffer, *TFR* Teflon without buffer, *TFBSR* Teflon without buffer in the presence of pieces of borosilicate, *ECP* elemental prebiotic chemical precursors.

Figure 3. Reaction products obtained in the experiments. **(A)** Comparison of molecular diversity among the three experiments. Note that amino acids, carboxylic acids, and nucleobases were always produced in the presence of borosilicate in total percentage higher than other products (C-1 chemical precursors and amines), thus favoring the mass balance towards the formation of compounds that are, in principle, useful intermediates for molecular evolution. **(B,C)** Optical micrographs of the wet **(B)** and dry **(C)** organic film covering the inner wall of the borosilicate flask; **(D)** infrared spectra of the organic skin; **(E)** Raman spectra of the organic skin after carbonation, showing D and G peaks. **(F)** EDX mapping of the organic film; **(G)** EDX elemental composition of the particle shown in **(H)** showing the existence of silicon in the film.

entry 4). A slightly different behavior was observed in the formation of aromatic miscellanea **41**–**48**, including polycyclic aromatic derivatives **45**–**48** (PAHs) (Fig. 2, Tables S2-S3). PAHs are important contributors to the overall pool of organic carbon in the universe and potential candidates in the "aromatic world" hypothesis[26]. Aromatic derivatives prevailed in the borosilicate flask under buffered conditions, but this trend was reversed in the absence of the buffer, in which case the highest total yield was obtained in Teflon alone (Table 1, entry 5). The effect of the buffer in the selectivity of the reactions and possible reaction pathways for the formation of compounds **1**–**48** are discussed in Supplementary text SI#5.

Discussion

Our results demonstrate that the wall of the reactors plays a crucial role in the synthesis of organic compounds in the Miller-Urey experiment. As summarized in Fig. 3A, the molecular diversity is minimal in the Teflon reactor, increases when submerging pieces of borosilicate glasses in the water of the Teflon reactor, and it reaches a maximum in both molecular variety and yielding in the borosilicate reactor. Furthermore, few hours after sparking, the wall of the borosilicate flask is covered by a thin brown film of organic matter. Noticeably, this film only forms on the part of the wall above the water level of the reactor. The color of the solution in the borosilicate reactors is yellow–brown and is full of brown organic particles visible to the naked eye. In none of the Teflon reactors, the formation of this organic film was observed. However, in the experiment performed with a Teflon reactor "seeded" with pieces of borosilicate glass, brown particles were noticed inside the solution.

The silanol groups on the surface of the glass, and traces of metal that could be released by dissolution under the alkaline conditions of the experiment may contribute to the observed reactivity[27,28]. The presence of Si–O–H groups enhanced by the alkaline conditions facilitates the absorption of the organic molecules synthesized in the gas and the liquid water in contact with the glass[29]. This could explain the formation of the brown film covering the inner surface of the borosilicate flask. The film appears as a translucent orange matrix under the optical microscope (Fig. 3B,C). The infrared and Raman spectra of the freshly formed film (Fig. 3D,E) show the characteristic absorption bands for HCN oligomers[30]. GC–MS confirms that the film is mainly made of HCN oligomers, in accordance with previously reported data. It also shows that it works as a matrix embedding and concentrating organic molecules, including urea **3**, glycine **5**, lactic acid **28**, adenine **36**, cytosine **39**, guanidine **49**, succinic acid **50**, 2,4-diamino-6-hydroxypyrimidine **51**, hypoxanthine **52**, and four polycyclic aromatic hydrocarbons, namely anthracene **53**, chrysene **54**, pyrene **55**, and dibenz(*a,h*) anthracene **56** (Fig. 2, Table S5). Among them, **49**–**56** were not previously detected in the liquid fraction of the experiment. As a general trend, the total yield of these latter compounds was found to increase after acid hydrolysis[31], highlighting the possibility that the treatment favored their extraction from the solid matrix (See supplementary information Table S5 condition A vs. condition B). The EDX analysis of the film reveals the existence of a significant amount of silica (Fig. 3F,H and Figure S8). The formation of organosilicon compounds is most likely responsible for the incomplete mass balance relative to the crude (Table 1). In addition, the highest total yield for the reaction products observed under unbuffered conditions is in accordance with a possible role of borosilicate as a catalyst for prebiotic processes (Table 1, entry 7).

From the initial bet of Bernal and Goldsmichdt for montmorillonite[32], many other minerals have been proposed to speed up the synthesis of specific molecules required for life as we know it, namely other clays, zeolites, sulfides, iron oxide, layered hydroxides, silica, etc.[15,33,34]. Experimental and theoretical work has been published to support these claims[35]. In particular, simple variations in environmental mineral composition lead to differentiation of distinct chemical pathways[36], encompassing the role of mineral surface in the prebiotic origin of amino acids[37] and peptides[38], mechanochemical solid-state transformations[39], and borosilicate-mediated formose condensation in the synthesis and stabilization of biologically relevant four and six-carbon sugars[38]. However, we still miss a good understanding of the structural reasons why and how mineral surfaces catalyze reactions relevant to prebiotic chemistry and the origin of life[40]. The importance of our results lies in the fact that, for the first time, the role of borosilicate has been experimentally demonstrated in a type of synthesis of the utmost relevance for the inorganic generation of organic compounds from scratch. The famous Miller-Urey synthesis triggered by sparking would be highly efficient at any place of the universe, provided a mineral surface is available. Noteworthy, silica and silicates also trigger the formation of insoluble organic matrices that serve as niches for the preservation and concentration of forming prebiotic molecules. These abiotic organic films may have formed in Earth-like planets and moons as Mars and several moons of the solar system[41–43]. For instance, a large fraction of the organic matter found in Archean rocks and to be found in the robotic exploration of Mars might reasonably be of inorganic origin. The putative role of the organic film triggered by the borosilicate reactors as a milieu for absorption and concentration of organic molecules should be further investigated. And indeed, the formation and properties of these organic films must be explored with different mineral surfaces and different atmospheres.

Conclusion

The experiment is especially important in the framework of the new ideas about the Hadean Earth in which the concomitance of a reduced atmosphere, electrical storms, silicate-rich rocky surfaces, and liquid water is expected[31,44]. Our results demonstrate that silica and silicates drastically enhance Miller's prebiotic synthetic routes affording important differences in the yields, in the number of products, and in their increased chemical information described by the number of carbon and nitrogen atoms composing the molecules, which are obtained starting from 1-carbon atom and 1-nitrogen atom precursors. Irrespective of the possible lack of correspondence[45,46] of the early Earth atmosphere with that originally proposed by Miller-Urey, these results show the efficiency and the prebiotic worth of the borosilicate/spark discharge system. The presence of high molecular weight products is exemplified by the presence of a dipeptide, of multi-carbon atoms dicarboxylic acids, of PAHs[47], of a complete panel of biological nucleobases, and, markedly, by the rich variety of different

classes of compounds. In summary, Miller recreated in his experiments the atmosphere and waters of the primitive Earth. The role of the rocks was hidden in the walls of the reactors.

Methods

The electric discharge was performed under unbuffered and buffered solution (NH_4Cl, 0.05 M, pH 8.7) in a Teflon apparatus and compared with a classical borosilicate reactor as a reference (these experiments will be indicated in the follow as TFR Teflon reactor unbuffered, TFRB Teflon buffered, BSR borosilicate reactor unbuffered and BSRB borosiclicate reactor buffered) (detailed experimental set-up is in SI#1). Two more experimental conditions were studied: (i) the electric discharge in the absence of the buffer; and (ii) the electric discharge in the Teflon apparatus in the presence of borosilicate bits (17 g), under both buffered (TFBSR/B) and unbuffered (TFBSR) conditions. After the work-up, the reaction was lyophilized and immediately analyzed by GC–MS. The samples stored at −80 °C for one or two weeks (to replicate analysis) showed the same composition of freshly analyzed counterparts. This control was performed in order to rule out possible ageing-related variations of the reaction products at −80 °C, a possibility that was previously highlighted[48]. In detail, in a round bottom flask *N,N*-bis-trimethylsilyl trifluoroacetamide (420 µL; Merck > 99%) and a solution of pyridine (200 µL; Merck > 99%) were added to 10 mg of crude of the reaction. The mixture was left under magnetic stirring at 90 °C for 4 h. Thereafter the solution was cooled down to 25 °C and 2.0 µL of the solution were used for the GC–MS analysis. Chromatographic conditions: CP8944 column (WCOT fused silica, film thickness 0.25 µm, stationary phase VF-5 ms, Øi 0.25 mm, length 30 m), injection temperature 280 °C, detector temperature 280 °C, gradient 100 °C × 2 min, then 10 °C/min for 60 min. GC–MS fragmentation spectra were recovered by using a triple quadrupole MS analyzer as full scan and single ion research modes, and compared with commercially available electron mass spectrum libraries. The libraries we used (NIST 2020 libraries; NIST/EPA/NIH Mass Spectral Library: c3oh_ci, c4h10_ci, ch4_all, ch4_drug, ch4_fda, libr_gp, libr_tr, libr_tx) are settled to contain more than 1.3 million spectra including most of the compounds of biological relevance and known products deriving from the chemistry of HCN and formamide[49]. These libraries also include isomeric structures. They tentatively identify unknown structure on the basis of the crossing of multiple experimental parameter values (i.e., retention time, m/z distribution and intensity of the corresponding fragmentation peaks)[50]. All products have been recognized with a similarity index (S.I.) greater than 98%. In the case of valine (**7**) isovaline (**17**), α-NH_2-isobutyric acid (**18**), γ-NH_2-butiric acid (**19**), butanamine (**23**) and isobutylamine (**24**), for which the similarity index was encompassed between 97 and 98%, the qualitative assignment was performed by co-injection method, repeating the GC–MS analysis after the addition of 0.1 µmol of appropriate standard compounds before the derivatization procedure (original co-injected chromatograms are in SI#3). The yield of reaction products was calculated in triplicate as micrograms of product per 1.0 mg of the crude and mg of product per total amount of the reaction crude, using the calibration line procedure, or in alternative (for compounds **17–19** and **23–24**) the internal standard method in the presence of betulinic acid (3β-hydroxy-20(29)-lupaene-oic acid) as internal standard (0.2 mg, 0.00045 mmol) (the general description of calibration line procedure and internal standard method is in SI#7).

Data availability

All data is available in the main text or the supplementary materials.

References

1. Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science* **117**, 528–529 (1953).
2. Ball, P. *Elegant Solutions: Ten Beautiful Experiments in Chemistry* (Royal Society of Chemistry, 2015).
3. Miller, S. L. Production of some organic compounds under possible primitive earth conditions1. *J. Am. Chem. Soc.* **77**, 2351–2361 (1955).
4. Oró, J. Synthesis of organic compounds by electric discharges. *Nature* **197**, 862–867 (1963).
5. Schlesinger, G. S. & Miller, S. L. Prebiotic synthesis in atmospheres containing CH_4, CO, and CO_2. *J. Mol. Evol.* **19**, 383–390 (1983).
6. Miyakawa, S., Yamanashi, H., Kobayashi, K., Cleaves, H. J. & Miller, S. L. Prebiotic synthesis from CO atmospheres: Implications for the origins of life. *PNAS* **99**, 14628–14631 (2002).
7. Cooper, G. J. T. *et al.* Miller-Urey spark-discharge experiments in the deuterium world. *Angew. Chem. Int. Ed.* **56**, 8079 (2017).
8. Wollrab, E. *et al.* Chemical analysis of a "Miller-type" complex prebiotic broth: Part I: Chemical diversity, oxygen and nitrogen based polymers. *Orig. Life Evol. Biosph.* **46**, 149–169 (2016).
9. Johnson, A. P. *et al.* The Miller volcanic spark discharge experiment. *Science* **322**, 404–404 (2008).
10. Parker, E. T. *et al.* Primordial synthesis of amines and amino acids in a 1958 Miller H2S-rich spark discharge experiment. *PNAS* **108**, 5526–5531 (2011).
11. Garcia-Ruiz, J. M. Self-assembled silica-carbonate structures and detection of ancient microfossils. *Science* **302**, 1194–1197 (2003).
12. Garcia-Ruiz, J. M., Melero-Garcia, E. & Hyde, S. T. Morphogenesis of self-assembled nanocrystalline materials of barium carbonate and silica. *Science* **323**, 362–365 (2009).
13. Montalti, M. *et al.* Local pH oscillations witness autocatalytic self-organization of biomorphic nanostructures. *Nat. Commun.* **8**, 14427 (2017).
14. Saladino, R. *et al.* Meteorite-catalyzed syntheses of nucleosides and of other prebiotic compounds from formamide under proton irradiation. *PNAS* **112**, E2746–E2755 (2015).
15. Saladino, R., Di Mauro, E. & García-Ruiz, J. M. A universal geochemical scenario for formamide condensation and prebiotic chemistry. *Chem. Eur. J.* **25**, 3181–3189 (2019).
16. Brandes, J. A. *et al.* Abiotic nitrogen reduction on the early Earth. *Nature* **395**, 365–367 (1998).
17. Zahnle, K., Schaefer, L. & Fegley, B. Earth's earliest atmospheres. *Cold Spring Harbor Perspect. Biol.* **2**, a004895 (2010).
18. Bada, J. L. & Miller, S. L. Ammonium ion concentration in the primitive ocean. *Science* **159**, 423–425 (1968).
19. Miller, S. L. & Van Trump, J. E. *Origin of Life*. 135–141 (Springer, 1981).
20. Saitta, A. M. & Saija, F. Miller experiments in atomistic computer simulations. *PNAS* **111**, 13768–13773 (2014).

21. Saladino, R., Botta, G., Pino, S., Costanzo, G. & Di Mauro, E. Genetics first or metabolism first? The formamide clue. *Chem. Soc. Rev.* **41**, 5526–5565 (2012).
22. Ferus, M. *et al.* Formation of nucleobases in a Miller–Urey reducing atmosphere. *PNAS* **114**, 4306–4311 (2017).
23. Ferus, M. *et al.* High-energy chemistry of formamide: A unified mechanism of nucleobase formation. *PNAS* **112**, 657–662 (2015).
24. Saladino, R., Barontini, M., Cossetti, C., Di Mauro, E. & Crestini, C. The effects of borate minerals on the synthesis of nucleic acid bases, amino acids and biogenic carboxylic acids from formamide. *Orig. Life Evol. Biosph.* **41**, 317–330 (2011).
25. Costanzo, G., Saladino, R., Crestini, C., Ciciriello, F. & Di Mauro, E. Formamide as the main building block in the origin of nucleic acids. *BMC Evol. Biol.* **7**, S1 (2007).
26. Ehrenfreund, P., Rasmussen, S., Cleaves, J. & Chen, L. Experimentally tracing the key steps in the origin of life: The aromatic world. *Astrobiology* **6**, 490–520 (2006).
27. Iler, R. K. *The Chemistry of Silica* (Wiley, 1979).
28. Rimola, A., Costa, D., Sodupe, M., Lambert, J. F. & Ugliengo, P. Silica surface features and their role in the adsorption of biomolecules: Computational modeling and experiments. *Chem. Rev.* **113**, 4216–4313 (2013).
29. Parida, S. K., Dash, S., Patel, S. & Mishra, B. K. Adsorption of organic molecules on silica surface. *Adv. Colloid Interface Sci.* **121**, 77–110 (2006).
30. Ruiz-Bermejo, M. *et al.* New insights into the characterization of 'insoluble black HCN polymers'. *Chem. Biodivers.* **9**, 25–40 (2012).
31. Zahnle, K. *et al.* Emergence of a habitable planet. *Space Sci. Rev.* **129**, 35–78 (2007).
32. Ferris, J. P. Montmorillonite-catalysed formation of RNA oligomers: the possible role of catalysis in the origins of life. *Philos. Trans. R. Soc. B* **361**, 1777 (2006).
33. Cairns-Smith, A. *Genetic Takeover and the Mineral Origins of Life* (Cambridge University Press, 1987).
34. Ferris, J. P. Prebiotic synthesis on minerals: Bridging the prebiotic and RNA worlds. *Biol. Bull.* **196**, 311–314 (1999).
35. Lambert, J. B., Senthil, A.G.-T. & Kuangbiao, M. The silicate-mediated formose reaction: Bottom-up synthesis of sugar silicates. *Science* **327**, 984–986 (2010).
36. Surman, A. J. *et al.* Environmental control programs the emergence of distinct functional ensembles from unconstrained chemical reactions. *PNAS* **116**, 5387–5392 (2019).
37. Vinogradoff, V. *et al.* Impact of phyllosilicates on amino acid formation under asteroidal conditions. *ACS Earth Space Chem.* **4**(8), 1398–1407 (2020).
38. Erastova, V. *et al.* Mineral surface chemistry control for origin of prebiotic peptides. *Nat. Commun.* **8**, 2033 (2017).
39. Haas, M. *et al.* Mineral-mediated carbohydrate synthesis by mechanical forces in a primordial geochemical setting. *Commun. Chem.* **3**, 14 (2020).
40. Hazen, R. M. & Sverjensky, D. A. Mineral surfaces, geochemical complexities, and the origins of life. *Cold Spring Harb. Perspect. Biol.* **2**, a002162 (2010).
41. Franz, H. B. *et al.* Indigenous and exogenous organics and surface–atmosphere cycling inferred from carbon and oxygen isotopes at Gale crater. *Nat. Astron.* **4**, 526–532 (2020).
42. Vago, J. L. *et al.* Habitability on early Mars and the search for biosignatures with the ExoMars Rover. *Astrobiology* **17**, 471–510 (2017).
43. Raulin, F., Brassé, C., Poch, O. & Coll, P. Prebiotic-like chemistry on Titan. *Chem. Soc. Rev.* **41**, 5380 (2012).
44. García-Ruiz, J. M., Van Zuilen, M. A. & Bach, W. Mineral self-organization on a lifeless planet. *Phys. Life Rev.* **34–35**, 62–80 (2020).
45. Trail, D., Watson, E. B. & Tailby, N. D. The oxidation state of Hadean magmas and implications for early Earth's atmosphere. *Nature* **480**, 79–82 (2011).
46. Cleaves, H. J., Chalmers, J. H., Lazcano, A., Miller, S. L. & Bada, J. L. A reassessment of prebiotic organic synthesis in neutral planetary atmospheres. *Orig. Life Evol. Biosph.* **38**, 105–115 (2008).
47. Bizzarri, B. M. *et al.* High-energy proton-beam-induced polymerization/oxygenation of hydroxynaphthalenes on meteorites and nitrogen transfer from urea: Modeling insoluble organic matter?. *Chemistry* **26**, 14919–14928 (2020).
48. Miyakawa, S., Cleaves, H. J. & Miller, S. L. The cold origin of life: A implications based on the hydrolytic stabilities of hydrogen cyanide and formamide. *Orig. Life Evol. Biosph.* **32**, 195–208 (2002).
49. Lai, Z. & Fiehn, O. Mass spectral fragmentation of trimethylsilylated small molecules. *Mass Spectrom. Rev.* **37**, 245–257 (2018).
50. Buch, A. *et al.* Development of a gas chromatography compatible sample processing system (SPS) for the in-situ analysis of refractory organic matter in martian soil: Preliminary results. *Adv. Space Res.* **43**, 143–151 (2009).

Acknowledgements

Author contributions

J.C.-R. and J.M.G.-R. conducted experiments and perform analysis. B.B. and R.S. perform organic chemistry analysis by GC–MS. J.M.G.-R., R.S. and E.D.M. wrote the text. J.M.G.-R. conceived the work. All authors discussed the results and agreed on conclusions.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-00235-4.

Correspondence and requests for materials should be addressed to J.M.G.-R. or R.S.

Reprints and permissions information is available at www.nature.com/reprints.