A need to integrate metagenomics and metabolomics in geosciences and develop the deep-time digital earth-biome database of India

Pradeep Srivastava, Prasanta Sanyal, Sharmila Bhattacharya, Praveen K. Mishra, Suryendu Dutta, Rajarshi Chakravarti, Niraj Rai, Naveen Navani, Anoop Ambili, K. P. Karanth, Jahanavi Joshi, Sushmita Singh and Senthil Kumar Sadasivam*

This article presents applications of metagenomics and metabolomics in geosciences. It emphasizes the significance of biomolecular proxies in palaeoclimatology, the evolution of life, the genesis of hydrocarbons and the role of biological processes in metallogeny. Several examples of breakthroughs with respect using these methods in earth sciences exist, such as the estimating resilience time of landscapes against invasive species. It is unfortunate that scientific programmes using bioproxies have not yet taken root in Indian institutions. Now is the appropriate time to delineate the critical role of biology in geology and establish it as a thrust area of research in India. A molecular geobiology programme would deal with the understanding of varied issues such as microbial heat production and its role in soil processes, the role of biology in mineralization, the use of biomarkers (metabolites) and ancient DNA studies in understanding feedbacks in climate change, evolution of life, etc. This article focuses on the use of metagenomics and metabolomics in palaeo-sciences and the potential intellectual dividends they could provide.

Keywords: Geosciences, metagenomics, metabolomics, metallogeny, palaeoclimatology.

Basic premise

THE geology of an area evolves by a combination of abiotic and biotic processes that synergistically drive each other and it is often considered that half of the geologic processes are biology-driven. However, the role of biology in geosciences is hardly recognized, except in the field of palaeontology. It holds promise in a multitude of paradigm shifts in the manner we understand and interpret geological records[1]. A good example is in the cycle of rock weathering in which lot is established about erosion and sedimentation, where climate and tectonics are considered the main drivers, but the role of microbes that modulates this process is poorly understood. Also, being controlled by the ambient chemical environment, fossil preservation records are often incomplete and therefore, what geologists observe in sedimentary sequences is an incomplete story. Thus, for example, in studies on the evolution of early life, the biogenicity (biological origin) of the fossils is often debated. In the current global warming scenarios, where biota and diversity are being increasingly compromised, a question that merits enquiry is the fate of surface geomorphic processes in an abiotic world[2,3]. Similarly, the microbial world that mediates processes ranging from silicate weathering, fixing and release of oxygen to metallogeny, though it plays a critical role does not get preserved as body fossils in rocks. Therefore, conventional geoscientific studies, as performed today in India, are not inclusive and require a new impetus that involves molecular biology. This article provides an overview of the global and national status of the application of molecular biology in geosciences and identifies scientific gaps that should be addressed.

Pradeep Srivastava, Rajarshi Chakravarti and Sushmita Singh are in the Department of Earth Sciences, Indian Institute of Technology, Roorkee 247 667, India; Prasanta Sanyal is in the Department of Earth Sciences, Indian Institute of Science Education and Research, Kolkata 741 246, India; Sharmila Bhattacharya and Anoop Ambili is in the Department of Earth and Environmental Sciences, Indian Institute of Science Education and Research, Mohali 140 306, India; Praveen K. Mishra is in the Department of Geology, School of Sciences, Cluster University of Jammu, Jammu 180 001, India; Suryendu Dutta is in the Department of Earth Sciences, Indian Institute of Technology Bombay, Mumbai 400 076, India; Niraj Rai is in the Birbal Sahni Institute of Palaeosciences, Lucknow 226 007, India; Naveen Navani is in the Department of Biosciences and Bioengineering, Indian Institute of Technology Roorkee 247 667, India; K. P. Karanth is in the Centre for Ecological Sciences, Indian Institute of Sciences, Bengaluru 560 012, India; Jahanavi Joshi is in the CSIR Center for Cellular and Molecular Biology, Hyderabad 500 007, India and Senthil Kumar Sadasivam is in the Geobiotechnology Laboratory/ PG and Research Department of Botany, National College (Autonomous), Tiruchirappalli 620 001, India

*For correspondence. (e-mail: pradeep@es.iitr.ac.in)

Metagenomics and metabolomics – definitions and scope

Metagenomics and metabolomics are subsets of molecular geobiology that track the microbial and biological changes preserved in sediments where physical fossil records are not available. Metagenomics is the study of genetic material extracted from sediments. Metabolomics uses biomarkers (like organic acids, proteins, amino acids and lipids) preserved in them to fingerprint the metabolic activity of the microbial community[4]. Biomarkers and ancient DNA (aDNA)/ palaeo environmental DNA (PalEnDNA) have helped address questions related to the emergence and evolution of life on Earth and biological feedbacks leading to warmer climates[5,6]. It is suggested that the conventional proxies used for reconstruction of the past climates may lose up to 40% of the crucial information on feedbacks that are biologically mediated and can be provided by understanding the microbes[7]. Metagenomics and metabolomics help buttress the classical data to refine 'geology-only'-based inferences on forcing, feedback, thresholds and time lag that are used in climate models. Further, debates on biogenicity and the syngenetic nature of fossil records can be resolved through the use of biomarkers[8]. The term 'biogenicity' reveals any morphological, chemical or mineralogical feature in rocks of all ages which are uniquely produced by organisms. Such fossil look-alike features may also otherwise have been produced by abiotic processes. The term 'syngenisis' implies that the fossil records and rocks in which they are preserved are time-synchronous.

Metagenomes (aDNA) get preserved in sediments and are extractable. Sequencing the extracted DNA provides an understanding of the palaeo biomes, but their preservation in sediments is limited to a few thousand years. This time interval, however, is sufficient to provide baseline data for the Anthropocene. Metabolites in sediments have a preservation potential of up to several millions of years and can be traced back to a particular organism and/or a habitat to longer timescales.

Geosciences research comprises studies of megafossils and identifiable microfossils, implying that only a fraction of the global biomass has been used to understand the evolution of life and related processes on Earth. It has been increasingly realized that microbes and other biological entities can mediate physico-chemical reactions through their metabolic activities. This realization that a large fraction of geological processes had been mediated through biological processes led to the new scientific discipline of 'biogeochemistry'.

Microbes are the first responders to climate shifts through the total transformation of certain communities into another, over timescales of weeks. Therefore, when preserved, the microbial remains provide an unprecedented high-resolution snapshot of the contemporary environment and hence the climate. In palaeoclimatology, proxies such as isotopes, elemental geochemistry, mineral magnetism, sediment characteristics and palynofacies are used. These have thresholds leading to variable response times for the same forcing and can vary from a few centuries to millennia[9]. The responses can often be nonlinear, determined by parameters that also change with time. Such proxies do not take cognizance of the role of microorganisms in mediating the responses and feedbacks of palaeoclimate proxies. Thus, to a large extent, palaeoclimate records are weakly constrained, making it difficult to develop sensible, process-based regional- and global-scale correlations.

Applications of metabolomics and metagenomics in geosciences: global scenario

Metabolomics

Application of biomarkers (metabolites) began with petroleum geology, where they were used to understand maturation and oil sources[10–12]. This field later gained momentum with the path-breaking work on palaeo reconstruction using biological markers by Seifert and Moldowan[13], who observed that oils from different geological settings have specific biomarkers. A Google search for keywords like 'biomarkers and petroleum geosciences', provided 11,400 articles which had been published, including 1000 in 2021. Understanding early life on earth is another frontier area of earth science that utilizes biomarkers. The evidence for the earliest life now has brought out enzyme-modulated signatures in rocks as old as 3.76 Ga. This discovery confirmed the biogenicity of earlier reported microfossils in West Greenland by Schopf[14]. Further lipid-based biomarkers were used as evidence of Archean life in the Pilbara craton, Australia[15]. According to Knoll[16] 'the Archean (>2500 Ma) observed that discovery now has put an end to the concepts that the Earth has remained a shadow land in which "probably" and "maybe" qualify nearly all paleobiological interpretations until now'.

Responses of microbes and larger fauna and flora to the Quaternary climate variability have been probed using biomarkers, beginning with the identification of fauna in the Quaternary sapropel (marine organic sediments) in the Eastern Mediterranean[17]. Attempts were made to track the variation in lipid biomarkers linked to variability in contemporary climates in lake sediments[18]. These initial publications impeded the application of biomarkers in palaeoclimate research. Globally around 7900 articles have been published till now which explore biomarkers in palaeoenvironmental reconstruction. They deal with a variety of issues such as: (i) Resilience and diversity of flora and fauna to climate variability; (ii) Human–livestock evolution, migration and dispersal, and the role of climate change; (iii) Domestication and dispersion of animals and food grains; (iv) Geological records of antibiotic resistance; (v) Modern and palaeo-biomes of polar lakes; (vi) Temporal changes in methanotrophs and methanogens capturing past redox conditions;

Table 1. Applications of metabolomics in studies on metallogenesis

Study area	Metallogenetic implications	Reference
Supergiant gold deposits of the Witwatersrand Basin, South Africa	Spatial association of gold and uranium with organic matter.	23
Polymetallic Pb–Zn–Ag deposit in Qixiashan, Nanjing, China	Gas chromatograph mass spectrometry revealed mobilization of Pb and Zn facilitated by organic fluids derived from bacteria and algae.	24
Gold deposits of the Carbon Leader Reef, Witwatersrand Basin, South Africa	Biomarkers and carbon isotopes along with the presence of acyclic isoprenoids pristane and phytane confirmed the biogenic origin of Au. Large-scale Au fixation was triggered by early photosynthetic bacteria (cyanobacteria?) either by oxidation or reduction in a reducing environment 2900 Ma ago.	25, 26
Experimental	The bacterium *Delftia acidovorans* responds to the presence of toxic and soluble gold by activating certain metabolic pathways that reduce ionic gold to solid nano-nuggets.	27
Australia	Soils overlying base metal, Au, Pt and U deposits were found associated with specific microbial communities and abundance of metal-resistant genes closely associated with the underlying deposit. This has led to the development of the BASE (Biomes of Australian Soil Environments) project containing phylogenetic and geochemical data of 1400 sites across Australia.	28–30
Stratabound Cu–Ag Lorena deposit, Chile	Pyrobitumen–fluid interaction was found to be the key driver for mineralization that provided a conducive redox trap for Cu sulphide precipitation.	31
Kupferschiefer Cu-deposit, Spremberg area, Germany	The δD composition of the *n*-alkanes and isoprenoids along with molecular oxygen content in the long-chain *n*-alkanes revealed that organic matter may preserve alteration haloes indicative of mineralization. The biomarkers also suggested a slightly oxidizing, Cu-rich hydrothermal fluid with an evaporitic origin as the primary source of mineralization.	32

(vii) Tracing the sources of sedimentary metabolites to precise flora and fauna; (viii) Human colonization: bacteria–human relationship; (ix) Palaeo-salinity and ocean acidification; (x) Arrival of exotic species; and (xi) extending back to the records of the Anthropocene.

Metallogeny

The association of organic matter with ore-minerals in low to moderate temperature (100°–350°C) ore deposits was recognized in the early 20th century[19]. By the 1950s, it was envisaged that biological sequestering of metals, sulphide accumulation by sulphate-reducing bacteria, adsorption of metals on organic colloidal particles and mobilization of metals by organometallic complexes are important pathways leading to metal accumulation in a variety of ore-forming environments[20–22]. These studies served as a nucleus for research on the role of organic matter in metallogeny and its possible use in exploration. Listed in Table 1 (refs 23–32) are noteworthy examples wherein organic matter/biomarkers have either been used directly as an exploration proxy or their association with certain ore deposits has led to the development of robust exploration models.

Metagenomes

Application of metagenomes started with the sequencing of fragmented DNA extracted from an extinct zebra species[33]. Thereafter, phylogenetic details of several extinct species of humans, animals, plants and microbes have been reconstructed overtime along with their response to varying environmental conditions[34,35]. Till recently, the focus was towards understanding human and archaeological remains[36,37] and palaeopathology[38]. In 2013, the complete genome of a half a million-year-old horse was sequenced[39]. Preservation of aDNA was an issue and subsequent research indicated that colder, high-latitude regions provided a better preservation environment but lower population density. The converse, poor preservation and higher population, exists in the equatorial regions[40]. Currently, a general consensus is that the best sequences in the tropics would be in sediments of <15 ka.

In studies on palaeoenvironment, the concept of holobiome of environments has been proposed, which takes cognizance of all DNAs in the sediments that are used to develop a snapshot of the contemporary biome, including microbes[41]. This became possible only in 2006 with next-generation DNA sequencing (NGS)[42]. NGS helped develop time-resolved evolution of genome sequences of palaeoenvironments[43,44]. Using conventional genome and NGS sequencing, more than 200 articles have been published that deal with understanding ancient human remains, plants, microbes, fungi, bacteria, viruses and larger mammals, and the response of these individual biotic elements to environmental changes[40]. About 20 articles every year utilize lake sediment DNA to provide a time series of complete terrestrial and aquatic holobiome to Late Pleistocene–Holocene climate and environmental change[45].

Other applications

Applications of aDNA also include understanding the critical zone, where rock-weathering processes mediated by endolithic microbes secreting enzymes are studied[46]. This entire field of microbe–rock interface is poorly understood and needs more research. Likewise, in the context of human impact on landscape, sedimentary DNA (DNA of organisms preserved in sediments) extracted from riverine sediments has been in China to track antibiotics routing in the terrestrial landscape and map the possible locales where antibiotic resistance could have developed[47]. It is assumed that the emergence of antibiotic-resistant microbes is a relatively modern phenomenon after antibiotic drugs were discovered. However, a study using aDNA on the remains of an elephant in 30-ka-old permafrost sediments revealed the presence of antibiotic-resistant microbes. This triggered a debate about whether antibiotic resistance is ancient and may have developed naturally also[48].

Studies on microbes in geological contexts have seen significant changes due to technological developments such as NGS (that analyses multiple genes simultaneously) and DNA microarray technology for functional genomics. Here, an array of genes from different microbial domains is constructed for analysing the functional diversity of microbial communities in a habitat, employed to analyse and explore the biogeochemical process and functional activities in the fields of agriculture, human health and ecosystem management. However, they have been exploited in a limited manner. The latest version of bacteria-based DNA microarray contains 394,984 genes for over 1500 functional gene families[49].

Research at the geosciences–molecular biology interface in India

In India, the application of biomarkers has recently gained some traction. However, the application of metagenomics remains to be used effectively. Biomarkers have been applied on a variety of geological issues in India, ranging from late Neoproterozoic–Early Cambrian life, Late Permian coal and Permian–Triassic carbonaceous mud rocks, Jurassic black shales, Cretaceous coal and condensates, Eocene lignites and crude oil, Oligocene, Miocene, Plio-Pleistocene ambers, and Quaternary successions. Table 2 provides details on various biomarkers and their interpretations in geosciences. Next, we provide an overview of major contributions from India.

Neoproterozoic–tertiary sequences

Biomarkers and carbon isotope data from Neoproterozoic–Early Cambrian crude oils and sediments provided information about deep-time palaeobiology and helped reconstruct the palaeowater column in an epicontinental sea in the western Indian subcontinent[50–52]. Biomarkers were used for understanding ecological catastrophes in the past and their import for the future, e.g. in a late Permian–Triassic succession[53], at the mass extinction boundary, suggesting deforestation of coal–forming vegetation and a decline in the vascular plant biodiversity. Studies on terpenoids from coal (Cretaceous) and lignite (Eocene) deposits provided information on the evolution of plant metabolites and cell functionality[54–57]. The presence of stromatolites and filamentous bacteria from the Archean rocks of Dharwar and the Eastern Ghats provides the earliest evidence of life, and these require a revisit using biomarkers[58,59].

Detection of nitrogen-bearing compounds that could be proteinaceous moieties in faunal remains such as Triassic coprolites, Cretaceous dinosaur eggshells and Eocene fish vertebrae, revealed excellent preservation of labile macromolecules in the deep time[60–62]. Further, studies on Late Cretaceous, Oligocene, Miocene and Plio-Pleistocene ambers in sediments revealed the presence of monoterpenoids in plants that indicated the emergence of defence mechanisms in the plants against herbivores[63–66].

Biomarkers and metalogeny

In India, occurrences of organic matter associated with ore deposits exist; but the identification of specific biomarkers in metallogeny has not been explored. For example, the prospect of gold across the North Singhbhum Mobile Belt in eastern India is closely associated with carbonaceous black shale, where organic matter played a role in gold accumulation during sedimentation and early diagenesis[67]. Kaur *et al.*[68] suggested using marine microbes for biosynthesis of Cu-nanoparticles. Using sequencing of ribosomal RNA, the bacterial strain *Kocuria flava* M-7 was identified as a potential means for ‘green synthesis’ of copper. Thus, our knowledge of biomarkers applied to metallogeny remains rudimentary and warrants an interdisciplinary approach involving organic geochemistry, molecular biology and economic geology.

Studies in the quaternary

Only a few studies from India have utilized biomarkers in the Quaternary sequences. They utilized biomarkers to track the source of organic matter in lakes, vegetation history and climate variability, biogenic signatures of water pollution, etc. Table 3 (refs 69–79) lists the notable Indian contributions in this regard.

It is known that in fluvial deposits, subsequent to deposition, the sedimentary organic matter (OM) undergo degradation with time and therefore may provide erroneous isotopic values for bulk organic sediments and lead to inaccurate interpretation of contemporary depositional environments ($\delta^{13}C_{\mathrm{bulk}}$)[4]. Therefore, a compound-specific isotopic analysis (CSIA) of a particular carbon (or hydrogen) chain length of *n*-alkane extracted from the sediments/vegetation was used to provide a more robust understanding of past vegetation ($\delta^{13}C_{\mathrm{CISA}}$) and moisture changes ($\delta$Dwax)[78,80]. CSIA also helped understand the effects of climate and

Table 2. Various biomarker compounds, their biotic sources and indicative environments

Compound	Biotic source	Environment
Short-chain normal alkanes (*n*-C_{15}, *n*-C_{17}, *n*-C_{19})	Microbes such as bacteria, algae	Marine, lacustrine
Mid-chain normal alkanes (*n*-C_{21}, *n*-C_{23}, *n*-C_{25})	Aquatic macrophytes, *Sphagnum* species	Lacustrine, environment mire
Long-chain normal alkanes (*n*-C_{27}, *n*-C_{29}, *n*-C_{31}, *n*-C_{33})	Terrestrial vascular woody trees, grasses	Terrestrial
Acyclic isoprenoids pristane (Pr) and phytane (Ph)	Phytol side chain of chlorophyll (bacterial, algal, terrestrial plants)	Pristane/phytane < 0.8 signifies highly anoxic condition
Saturated diterpanes (e.g. phyllocladane, isopimarane, abietane); aromatic diterpanes (retene, simonellite)	Gymnosperms	Terrestrial
C_{19}–C_{33} tricyclic terpanes	Orphan biomarkers; considered to be derived from algae or bacteria	Mostly marine
C_{24} tetracyclic terpanes	Unknown	Hypersaline
C_{27}-C_{29} steranes	Algae, terrestrial vascular plants. In the Precambrian, C_{27} was largely derived from red algae and C_{29} from green algae. In the Phanerozoic, C_{29} was from vascular plants	Marine, lacustrine, terrestrial
C_{30} 24-*n*-propylcholestane	Chrysophyte algae, foraminifera	Marine
C_{30} 24-*iso*-propylcholestane	Demosponges	Marine (mostly)
Pregnane, homopregnane	Unknown	Hypersaline
4-Methylsteranes	Algae, dinoflagellates	Marine, lacustrine
Dinosterane	Dinoflagellates	Marine (mostly), younger than Triassic (some exceptional Precambrian occurrences)
Diasteranes	Algae, vascular plants	Formed through rearrangement reactions in clay-rich rocks
Lanostane	Eukaryotes (multiple sources)	Various
Secosteranes	Unknown	Various, biodegradation?
C_{27} trisnorhopane (Ts and Tm)	Bacteria	Marine, terrestrial, Ts increases with increasing thermal maturity
C_{28} Bisnorhopane	Bacteria	Unusual occurrence
C_{29} 30-norhopane	Bacteria	Carbonate/evaporites
C_{29} 25-norhopane	Microbial reworking	Likely formed during biodegradation
C_{30} hopane	Bacteria	Marine, terrestrial
C_{31}–C_{35} homohopanes	Bacteria	Marine, terrestrial, C_{35} homohopane concentration is higher in reducing/anoxic condition
2α-Methylhopane	Cyanobacteria	Marine, lacustrine
3β-Methylhopane	Methanotrophic bacteria	Various
Gammacerane	Ciliate protozoan (such as *Tetrahymena*)	Stratified water column induced by salinity difference
Oleanane	Angiosperms	Terrestrial
Bicadinane	Angiosperms	Terrestrial, dominant Dipterocarpaceae input
Carotenoids (e.g. chlorobactane, okenane, isorenieratane)	Sulphur bacteria (purple sulphur bacteria, green-pigmented green sulphur bacteria, brown-pigmented green sulphur bacteria)	Photic zone euxinia
Polycyclic aromatic hydrocarbons (especially in surface or younger sediments)	Aromatization of various compounds in sediments	Industrial/domestic waste in various ecosystems, oil spillage (including saturated petrogenic hydrocarbons), many are carcinogenic
Coprostanol	Conversion of cholesterol by microbial hydrogenation in mammalian guts and hence detected in the faeces of mammals	Indicator of sewage pollution or usage of manure as fertilizer (agricultural practices)
Bile acids	Digestive tract of animals	Indicator of faecal matter

associated physico-chemical processes on the fate of plant-derived OM in modern soils and paleosols[79,81–84]. Similar attempts have been made to understand the depositional and vegetation history of the Siwalik sedimentary rocks of India. Table 4 (refs 85–87) provides a brief summary of such contributions.

Several centres in India (e.g. IIT Bombay, IISc, IISER Kolkata, IISER Mohali and BSIP) have started using

Table 3. Application of biomarkers in Quaternary sediments in various lakes and fluvial records of India

Archives	Location	Proxies	Remarks	Reference
Lake	Lonar Lake (central India)	*n*-Alkane chain length, and ACL	To understand the factors controlling the spatial distribution of organic proxies within the Lake	69
	Ennamangalam Lake (Tamil Nadu, South India)	CSIA ($\delta^{13}C_{n\text{-alkane}}$), alkane chain length indices (TAR, CPI)	Using lipid biomarkers (*n*-alkanes) distribution and CSIA of carbon ($\delta^{13}C_{n\text{-alkane}}$), work on sediments and modern vegetation samples in and around Ennamangalam Lake revealed past climatic variability in response to ISM variability during mid–late Holocene	70
	Beni Tal (Garhwal Himalaya)	CSIA (δD_{C29}, and $\delta^{13}C_{C29}$)	Records early Holocene ISM and its impact on vegetation in the Central Himalaya	71
	Ahansar Lake (Kashmir Himalaya)	Organic proxies (e.g. $\delta^{15}N$; amino acid indices, e.g. RI, DI, Ox/Anox ratio), and pollen assemblages	Provides long-term factors controlling natural as well as anthropogenic components in a lake basin	72
	Ahansar Lake (Kashmir Himalaya)	Along with organic proxies, the *n*-alkane chain length (e.g. monomethyl alkanes – MMAs; highly branched isoprenoids – HBIs, terpenoids and Pristane/ Phytane ratio)	Proxy helps decipher the factors controlling the apportioning of sedimentary organic matter sources and their degradation state in the anthropogenically impacted lake basins.	73
Peat deposits	Kedarnath (Garhwal Himalaya)	CSIA (δD_{C29} and $\delta^{13}C_{C29}$), and *n*-alkane proxies (e.g. CPI, ACL)	The organic proxies help decipher the reconstruction of past ISM vegetational variability in Garhwal Himalaya	74
Sediment profile	Dzokou Valley (Northeastern Himalaya)	*n*-Alkane chain length and various *n*-alkane indices (CPI, P_{aq})	Records vegetational changes in North East India during the Late Holocene in terms of ISM variability in the region	75
	Banni Grassland (western India)	CSIA ($\delta D_{n\text{-alkane}}$, $\delta^{13}C_{n\text{-alkane}}$), TAR, CPI, P_{aq}	Impact of monsoonal precipitation during the mid–late Holocene on grassland ecosystem of the region	76
	Ganga Plain	*n*-Alkanoic acids, *n*-alkane, and $\delta^{13}C_{\text{fattyacids}}$	Reveals past vegetational changes in the Ganga Plain during the last glacial cycle	77
	Belan River (north-central India)	Soil carbonate ($\delta^{13}C$, $\delta^{18}O$), CSIA (δD_{C29} and $\delta^{13}C_{C29}$ of paleosol)	Utilization of proxies for rainfall ($\delta^{13}C$, $\delta^{18}O$) and vegetation change ($\delta^{13}C_{C29}$ and $\delta^{13}C_{SC}$) on Quaternary fluvial sequences of Belan River in north–central India revealed the factors (such as climatic fluctuation and vegetational changes) influencing human settlement in the region	78
	Mahi River (western India, Thar Desert)	$\delta^{13}C_{pc}$ and $\delta^{18}O_{pc}$ of paedogenic carbonate, $\delta^{13}C$ of bulk organic matter and *n*-alkane	Utilizing these proxies, the effect of sub-aerial exposure and pedogenesis on calcrete using C and O isotopic composition was recorded. Further, the study indicated the usefulness of CSIA, where $\delta^{13}C_{n\text{-alkane}}$ values of the sediment profiles were unaffected by the pedogenic modification, thus providing a better picture of the past vegetational condition	79

biomarkers to understand natural Earth processes in terms of changes in the past climate and vegetation, the depositional environment, metallogenesis, causes of ecological catastrophes and the extent of anthropogenic impact on natural systems[88–96].

Environmental pollution

In aquatic systems, the source apportionment of OM (authigenic versus terrestrial and anthropogenic) is crucial to understand the carbon cycle. However, increasing population load around the waterbodies negatively impacts their natural ecosystem and distorts the natural carbon cycle. Therefore, studies on aquatic systems using biomarkers offer an opportunity to improve our understanding of the organic contaminants. Bulbul *et al.*[97]. utilized *n*-alkane indices, the ratio of isoprenoid compounds and hopanes to characterize the human activity and petroleum contamination in the Mandovi estuary, Goa (west coast of India). Increasing terrigenous/aquatic ratio (TAR) (and decreasing *P*-aqueous (P_{aq})) ratios indicate the dominant contribution of terrigenous OM from the upper estuary. In addition, the Pr/Ph

Table 4. Application of biomarkers in Indian Siwaliks

Archives	Location	Proxies	Remarks	Reference
Sediment profile	Himalayan foreland basin (Tista Valley, NE Himalaya)	CSIA ($\delta^{13}C_{n\text{-alkane}}$), alkane chain length indices (e.g. sterane/17α-hopane, $C_{31}R/C_{30}$-hopane and C_{27}/C_{29} steranes ratios)	Revealed the influence of marine incursion during 3–10 Myr in the Tista Valley sequence which was supposed to be exclusively fluvial-dominated sediments	85
	Siwalik paleosols (NW Himalaya)	CSIA (δDC_{29} and $\delta^{13}C_{C29}$), n-alkane based indices (e.g. CPI, ACL)	These proxies help decipher the evolution of C4 plants and their controlling factors (such as variability in ISM rainfall, type of overbank sediments and moisture content) in the Siwalik Himalaya	86
	Siwalik paleosols (Himachal Pradesh)	Inorganic (REE and trace elements) and organic proxies (such as Pristane/Phytane, D/H, $\delta^{13}C_{n\text{-alkane}}$)	Alluvial fan deposits of the Siwaliks from Kangra sub-basin have been used to study the combined impact of depositional environment and climatic and vegetational conditions on paleosol formation.	87

ratio (pristane/phytane) was used to identify the extent of petroleum contamination in the estuary closer to the sea (lower estuary) and the relative degradation of n-alkanes[97].

Biomarkers also provide a crucial understanding of environmental pollution. Anthropogenically derived organic pollutants or synthetic polymers (such as phthalates) that may enter the aqueous or soil systems prove to be detrimental to the ecosystems. Hence, quantitative data are being generated from various lake-system estuaries in India to assess the extent of pollution in the waterbodies[97,98]. These studies have provided baseline information for future research on the impact of hazardous materials in diverse ecosystems.

Microplastics are a serious threat to human health and aquatic ecosystems. In India, studies are being carried out to understand the spatial distribution as well as the composition of microplastics in the aquatic systems. However, only limited data are available from the Himalayan region. Ajay *et al.*[98] studied the microplastics and phthalic acid esters (PAEs) (such as di-isobutyl phthalate, diethyl phthalate, di-(2-ethylhexyl) phthalate, etc.) in the Renuka Lake, Sirmaur district, to understand the extent of anthropogenic activity in Himachal Pradesh. The results indicated that phthalates could be used quantitatively to represent microplastic pollution in aquatic systems[99–101].

Metagenomics

Currently, our knowledge of the microbial population in freshwater terrestrial sediments is meagre. A preliminary study conducted in India on sediments from an Antarctic lake demonstrated the applicability of aDNA/PalEnDNA in lake sediments[102]. A recent study on the sedimentary core of the Kaveri Delta tracked sea-level changes using microbes[103]. A study tracked the genetics of ostriches in India during the Late Quaternary[104]. Except for a few publications on phylogenetic analysis of partially sequenced 16S rRNA gene sequences from the Sundarban sediments[105], and analysis of bacterial community structure and function in sediment ecosystems of the Bay of Bengal[105], the genomic data in India are scarce and microbial responses to climate change have not been examined.

In the context of archeo-geobiology, the spatial and temporal response of human populations (e.g. evolution, diversification and migration) to climate change has been of interest for a long time. Molecular studies using aDNA are poorly understood in the Indian context and only a few archaeological sites like Rakhi Garhi, Roopkund and Ajnala have been explored[106–108].

This is an opportune time to initiate aDNA research on geological samples to reconstruct the past climatic conditions in space and time. Besides providing past environmental data, aDNA data of the extant and extinct species can be modelled to predict the future consequences of climate.

Questions to be addressed in the Indian context: quaternary and pre-Quaternary

Quaternary witnessed several glacial and interglacial cycles with overriding events like the Early Holocene Climatic Optimum and Medieval Climatic Anomaly, when climate reversed from warm and wet to cold and dry, and vice versa within a few thousand years. In the current global warming scenario, the trend of temperature changes is similar to that witnessed during the Quaternary, though the rates of change now are much faster. A large part of the globe witnessed megafaunal extinction associated with warming and the spreading of humans during the Pleistocene. The role of warming-related outbreaks of bacteria and viruses has not been fully explored. In a global warming scenario where environmental temperatures are projected to rise by 1.5°C by 2050 with increased frequency of extreme weather events[109,110], increased glacial melt and higher erosion are anticipated. Such past events can provide a model of microbiological changes that accompany climate reversals.

In the context of the Anthropocene, problems that can be examined using metagenomics and metabolomics are as follows

- (i) What would be the magnitude of a turnaround in the microbial world, and what lessons can be learnt from the past?
- (ii) What will the microbial world look like in 2050 in the scenario of global warming? Which new species in microbiomes will or can emerge because of climate change?
- (iii) Do metagenomes and biomarkers have the ability to provide early warnings of extreme climatic events and global warming?
- (iv) How will the change in biota affect the soils and geomorphic processes?
- (v) What will be the nature and resilience of pathogens and what will be the implications on the development of antibiotic resistance? Can we have clues on natural antibiotic resistance in sedimentary archives?
- (vi) How have microbiomes of lakes, wetlands and permafrost responded to climatic warming and transitions in the past?
- (vii) Are climate and tectonics only responsible for weathering and mass-wasting processes in the Himalayas and Deccan plateau region?
- (viii) What are the causes of the Pleistocene mammalian extinction?
- (ix) Can there be any biological precursors to forest fires? What does palaeoclimate record teach us?
- (x) Genome mapping of environmental DNA (eDNA).

These questions can only be answered through integrating geosciences and microbiology in a study of time-resolved biotic responses to changes in the Earth system. To start with, records of microbiotic responses through geological time, including the periods of climatic extremes and mass extinctions, are needed to serve as a template for the present[111].

In the pre-Quaternary timescales, there are complex issues related to the evolution of life, where the biogenicity of fossil records is often debated. There are major oxygenation and hypoxia events that must be understood in terms of biotic responses and feedbacks. Some of the important issues/questions are the following

- (i) How did life evolve on Earth and how Biogenecity and syngenecity of fossils should be established?
- (ii) Past records of ocean acidification (increasing atmospheric carbon dioxide will eventually acidify the oceans).
- (iii) In the context of Banded Iron Formation (BIF), what led to the early oxidation of the Archean oceans in India?
- (iv) What led to the Archean glaciation?
- (v) What was the cause of the biological explosion at the Precambrian–Cambrian boundary?
- (vi) How and which part of metallogeny and biomineralization was mediated by the microbes?

What is needed?

This article surveys the use of biomolecular proxies in geosciences and the progress therein over the past few decades. India's geological richness includes sequences with signatures of the evolution of early life; oxygenation events and formation of BIFs; metallogeny; mass extinctions; faunal and floral evolution vis-à-vis growth of the Himalayas; wetlands, evolving human societies that offers incentives to add biological analysis to the molecular level and to refine our understanding of past ecosystems and their future trajectories in a warmer world. This calls for a database comprising geology, aDNA and biomarkers related to climate, life, ecosystem and metallogeny.

Synergy

In India, some of the centres mentioned earlier have functional groups and laboratories for molecular biogeosciences. Even in these institutions, efforts have been constrained due to limited financial support and human resources. This calls for a concerted effort from funding agencies and institutions to ensure that metabolomic and metagenomic applications grow as a research area in geosciences in the country.

Resources

Geological field-work, identification of research problem-based sections and strategic sampling in sterilized conditions (for aDNA and biomarkers) are the first steps in this field of research. The second step requires a clean and sterilized laboratory with RTP-CR, DNA extraction and microbial culture facilities, while the third step requires sequencing the extracted DNA. The second and third steps need funds to tune at least one lakh rupees per sample. Generating a time series from a geological section would require the analysis of several samples for a robust analysis. Large databases are the key and would need synergies and prioritization among individuals, institutions and funding agencies. To place the efforts needed in the context, some important initiatives are: (a) the 2018 Earth Biogenome Project (EBP) that aims to map the whole of Darwin's tree of life in the United Kingdom has a cost estimate of 4.7 billion USD (shared between UK, USA and China[112]). (b) Earth Microbiome Project (EMP) to generate a catalogue of environmental DNA of 200,000 globally distributed species from various ecosystems[113]. EMP received funding jointly from the John Templeton Foundation, W. M. Keck Foundation, Argonne National Laboratory, Australian Research Council, Tula Foundation and the Samuel Lawrence Foundation. (c) The Biomes of Australian Soil Environment (BASE) project in Australia will map the continent's soil biodiversity and the relationship between environmental parameters and biodiversity[29]. BASE was funded by a consortium of five funding agencies in Australia that included the Australian Government National Collaborative Research Infrastructure Strategy (NCRIS) and the Commonwealth Scientific and Industrial Research Organization (CSIRO). The Department of Biotechnology, Government of India has also launched a national programme, with 20 research organizations on-board, called 'Genome India Project (GIP)' that aims to study predictive diagnostic genome indicators of high-priority diseases. GIP carries a budget of Rs 238 crores (source: https://www.vifindia.org/2020/february/25/understanding-indias-genome-project). No programme in India can match the large-scale geosciences-based programmes like EBP, EMP and BASE. There is no denying that such a programme is needed for future planning, and an ecosystem accordingly has to be incubated and nurtured. Due to the enormity of resources magnified, such a programme can be conducted in phases as outlined below.

Phase-I can focus on the identification of thrust areas of research, laboratories and workforce augmentation and publishing the key research agenda. Phase-II can consider developing centres of excellence and laboratories, and introduce molecular biogeosciences in the undergraduate and postgraduate curricula. Phase-III should out-target nationwide programmes like EMP.

Deliverables of such a programme will be the generation of: (i) environmental indicators of climate change, pollution, invasive species and landscape degradation; (ii) geological records of holobiome evolution and the results of global warming; (iii) empowered laboratories and workforce; (iv) intellectual property in the form of Earth-bound genome catalogue of India; (v) newer avenues of job creation and (vi) to develop deep-time digital earth-biome database of India, in accordance with the Deep-time Digital Earth (DDE) programme of IUGS.

Limitations and challenges will be the poor preservation of aDNA due to climate variation and increased humidity and temperature in tropical and subtropical conditions[113]. Siwalik and Narmada Valley fossils, mammals and Quaternary fossils, including primates belonging to the Kurnool caves, also hold great potential and may yield some exciting results about the migration, evolution and phylogenetic relationships; however, the preservation limits of DNA may hinder such explorations. Collagen proteins have higher chances of preservation and have recently been used in evolutionary and phylogenetic analysis (for example, the phylogeny of the giant extinct ape *Gigantopithecus*)[114]. Further, improvement in techniques like the single- and double-stranded DNA library and in sequencing technology like high-throughput sequencing that requires DNA in very small amounts can change the scenario[41]. There is great potential to study the climate–society–biome relationship in the Himalayan regions, where temperatures are cooler and will favour DNA. In the hotter Ganga Plain, relatively larger amounts of aDNA preservation may compensate for the climate-related limitations.

There is a need for geologists and biotechnologists to further explore aDNA and how technology can combat the problem of low preservation.

Conclusion and recommendations

The literature survey indicates that the application of molecular biology and 'omic' technologies can play a significant role in tracking: (i) the evolution of biota and landscape at various times from the Archean to the Anthropocene; (ii) the feedback that microbes and biology at large provide under changing climate conditions and (iii) biological feedbacks in metallogeny and crustal evolution. In the Indian context, there has been some progress as far as the application of metabolomics is concerned, but metagenomics and/or functional genomics in geosciences have not been understood.

In the scenario of global warming, how the microbial world will respond can only be tracked by understanding how it did in the past. Such a study may link biological evolution and identification of genes that help adapt to modifications that the Earth system undergoes due to environmental changes. Therefore, this thrust area needs to be developed and will have immense societal benefits towards one of the most glaring dangers that the world is facing today.

---

1. Ficetola, G. F. *et al.*, DNA from lake sediments reveals long-term ecosystem changes after a biological invasion. *Sci. Adv.*, 2018, **4**(5), 4292.
2. National Research Council, *Landscapes on the Edge: New Horizons for Research on Earth's Surface*, The National Academies Press, Washington, DC, USA, 2010.
3. Corenblit, D. *et al.*, Feedbacks between geomorphology and biota controlling Earth surface processes and landforms: a review of foundation concepts and current understandings. *Earth-Sci. Rev.*, 2011, **106**(3–4), 307–331.
4. Eglinton, T. I. and Eglinton, G., Molecular proxies for paleoclimatology. *Earth Planet. Sci. Lett.*, 2008, **275**, 1–16.
5. Cooper, A. *et al.*, Abrupt warming events drove Late Pleistocene Holarctic megafaunal turnover. *Science*, 2015, **349**(6248), 602–606.
6. Singh, B. K. *et al.*, Microorganisms and climate change: terrestrial feedbacks and mitigation options. *Nature Rev. Microbiol.*, 2010, **8**(11), 779–790.
7. Rawlence, N. J. *et al.*, Using palaeoenvironmental DNA to reconstruct past environments: progress and prospects. *J. Quarter. Sci.*, 2014, **29**(7), 610–626.
8. Wang, M. *et al.*, A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabolism and planet oxygenation. *Mol. Biol. Evol.*, 2011, **28**(1), 567–582.
9. Singhvi, A. K. and Kale, V. S., *Paleoclimate Studies in India: Last Ice Age to the Present*, Indian National Science Academy, New Delhi, 2010.
10. Barker, C., Organic geochemistry as a geologic tool. *AAPG Bull.*, 1981, **65**(5), 894.
11. Reed, W. E., Molecular compositions of weathered petroleum and comparison with its possible source. *Geochim. Cosmochim. Acta*, 1977, **41**(2), 237–247.

12. Seifert, W. K. and Moldowan, J. M., Applications of steranes, terpanes and monoaromatics to the maturation, migration and source of crude oils. *Geochim. Cosmochim. Acta*, 1978, **42**(1), 77–95.
13. Seifert, W. K. and Moldowan, J. M., Paleoreconstruction by biological markers. *Geochim. Cosmochim. Acta*, 1981, **45**, 783–794.
14. Schopf, J. W., New evidence of the antiquity of life. *Origin. Life Evol. Biosphere*, 1994, **24**(2), 263–282.
15. Brocks, J. J. *et al*., Archean molecular fossils and the early rise of eukaryotes. *Science*, 1999, **285**(5430), 1033–1036.
16. Knoll, A. H., A new molecular window on early life. *Science*, 1999, **285**(5430), 1025–1026.
17. Smith, D. J. *et al*., Occurrence of long-chain alkan-diols and alkan-15-one-1-ols in a Quaternary sapropel from the Eastern Mediterranean. *Lipids*, 1983, **18**(12), 902–905.
18. Meyers, P. A. and Benson, L. V., Sedimentary biomarker and isotopic indicators of the paleoclimatic history of the Walker Lake basin, western Nevada. *Org. Geochem.*, 1988, **13**(4–6), 807–813.
19. Siebenthal, C. E., Origin of the lead and zinc deposits of the Joplin region. *US Geol. Surv. Bull.*, 1915, **606**, 283.
20. Berger, W., The geochemical role of organisms. *Mineral. Petrol. Mitteilung*, 1950, **2**, 136–140.
21. Barton, P. B., Possible role of organic matter in the precipitation of the Mississippi Valley ores. *Econ. Geol. Monogr*., 1967, **3**, 371–378.
22. Skinner, B. J., Precipitation of Mississippi Valley-type ores: a possible mechanism. *Econ. Geol. Monogr*., 1967, **3**, 363–370.
23. Haering, T. C., Organic geochemistry of Precambrian rocks. In *Researches in Geochemistry* (ed. Abelson, P. H.), Wiley, New York, USA, 1967, vol. 2, pp. 287–111.
24. Shucheng, X. *et al*., Biomarkers in fluid inclusions of polymetallic deposit of Qixiashan, Nanjing. *Chin. Sci. Bull.*, 1967, **42**(14), 1206–1209.
25. Spangenberg, J. E. and Frimmel, H. E., Basin-internal derivation of hydrocarbons in the Witwatersrand Basin, South Africa: evidence from bulk and molecular δ^{13}C data. *Chem. Geol.*, 2001, **173**(4), 339–355.
26. Frimmel, H. E. and Hennigh, Q., First whiffs of atmospheric oxygen triggered onset of crustal gold cycle. *Miner. Deposita*, 2015, **50**, 5–23.
27. Johnston, C. W. *et al*., Gold biomineralization by a metallophore from a gold-associated microbe. *Nature Chem. Biol.*, 2013, **9**, 241.
28. Reith, F. *et al*., Geogenic factors as drivers of microbial community diversity in soils overlying polymetallic deposits. *Appl. Environ. Microbiol.*, 2015, **81**, 7822–7832.
29. Bissett, A. *et al*., Introducing BASE – the Biomes of Australian soil environments soil microbial diversity database. *GigaScience*, 2016, **5**, 10–21.
30. Sanyal, S. S., Shuster, J. and Reith, F., Cycling of biogenic elements drives biogeochemical gold cycling. *Earth Sci. Rev.*, 2019, **190**, 131–147.
31. Herazo, A. *et al*., Assessing the role of bitumen in the formation of stratabound Cu-(Ag) deposits: insights from the Lorena deposit, Las Luces district, northern Chile. *Ore Geol. Rev.*, 2020, **124**, 103639.
32. Poetz, S. *et al*., Assessing the role of bitumen in the formation of stratabound Cu–(Ag) deposits: insights from the Lorena deposit, Las Luces district, northern Chile. *Org. Geochem.*, 2022, 1**69**, 104421.
33. Higuchi, R. *et al*., DNA sequences from the quagga, an extinct member of the horse family. *Nature*, 1984, **312**, 282–284.
34. Nielsen-Marsh, C., Biomolecules in fossil remains: multidisciplinary approach to endurance. *Biochemist*, 2002, **24**, 12–14.
35. Pääbo, S., Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc. Natl. Acad. Sci. USA*, 1989, **86**(6), 1939–1943.
36. Pääbo, S. *et al*., Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acids Res.*, 1988, **16**(20), 9775–9787.
37. Green, R. E. *et al*., A draft sequence of the Neanderthal genome. *Science*, 2010, **328**, 710–722.
38. Donoghue, H. D., Insights gained from ancient biomolecules into past and present tuberculosis – a personal perspective. *Int. J. Infect. Dis.*, 2017, **56**, 176–180.
39. Orlando, L. *et al*., Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, 2013, **499**, 74–78.
40. Parducci, L. *et al*., Ancient plant DNA in lake sediments. *New Phytol.*, 2017, **214**(3), 24–42.
41. Orlando, L. *et al*., Ancient DNA analysis. *Nature Rev. Methods Primers*, 2021, **1**(1), 1–26.
42. Goodwin, S. *et al*., Coming of age: ten years of next-generation sequencing technologies. *Nature Rev. Genet.*, 2016, **17**, 333–351.
43. Willerslev, E. *et al*., Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, 2014, **506**, 47–51.
44. Thomas, S. P. *et al*., Legacy of a Pleistocene bacterial community: patterns in community dynamics through changing ecosystems. *Geomicrobiol. J*., 2019, **35**(9), 798–803.
45. Capo, E. *et al*., Lake sedimentary DNA research on past terrestrial and aquatic biodiversity: overview and recommendations. *Quaternary*, 2021, **4**(1), 6.
46. Walker, J. J. and Pace, N. R., Endolithic microbial ecosystems. *Annu. Rev. Microbiol.*, 2007, **61**, 331–347.
47. Pei, R. *et al.,* Effect of river landscape on the sediment concentrations of antibiotics and corresponding antibiotic resistance genes (ARG). *Water Res.*, 2006, **40**(12), 2427–2435.
48. D'Costa, V. M. *et al*., Antibiotic resistance is ancient. *Nature*, 2011, **477**(7365), 457–461.
49. Li, H. *et al*., Responses of soil microbial functional genes to global changes are indirectly influenced by aboveground plant biomass variation. *Soil Biol. Biochem*., 2017, **104**, 18–29.
50. Dutta, S. *et al*., Biomarker signatures from Neoproterozoic–Early Cambrian oil, western India. *Org. Geochem.*, 2013, **56**, 68–80.
51. Bhattacharya, S. and Dutta, S., Neoproterozoic–Early Cambrian biota and ancient niche: a synthesis from molecular markers and palynomorphs from Bikaner–Nagaur Basin, western India. *Precambrian Res.*, 2015, **266**, 361–374.
52. Bhattacharya, S. *et al*., A distinctive biomarker assemblage in an Infracambrian oil and source rock from western India: molecular signatures of eukaryotic sterols and prokaryotic carotenoids. *Precambrian Res*., 2017, **290**, 101–112.
53. Bhattacharya, S. *et al*., Biotic response to environmental shift during the late Permian–Early Triassic transition: assessment from organic geochemical proxies and palynomorphs in terrestrial sediments from Raniganj Sub-basin, India. *Palaeogeogr.*, *Palaeoclimatol.*, *Palaeoecol*., 2021, **576**, 110483.
54. Mallick, M. *et al*., Pyrolytic and spectroscopic studies of Eocene resin from Vastan lignite mine, Cambay Basin, western India. *J. Geol. Soc. India*, 2009, **74**, 16–22.
55. Dutta, S. *et al*., Petrology, palynology and organic geochemistry of Eocene lignite of Matanomadh, Kutch Basin, western India: implications to depositional environment and hydrocarbon source potential. *Int. J. Coal Geol.*, 2011, **81**, 91–102.
56. Mathews, R. P. *et al*., Palynology, palaeoecology and palaeodepositional environment of Eocene lignites and associated sediments from Matanomadh Mine, Kutch Basin, Western India. *J. Geol. Soc. India*, 2013, **82**, 236–248.
57. Paul, S. and Dutta, S., Biomarker signatures of Early Cretaceous coal-bearing sediments of Kutch Basin, western India. *Curr. Sci*., 2015, **108**, 211–217.
58. Venkatachala, B. S. *et al*., Archaean microbiota from the Donimalai formation, Dharwar supergroup, India. *Precambr. Res*., 1990, **47**(1–2), 27–34.
59. Das Sharma, S. and Srinivasan, R., Stable isotope evidence for ca. 2.7-Ga-old Archean cap carbonates from the Dharwar Supergroup, southern India. *Curr. Sci.*, 2015, **108**(12), 2223–2229.

60. Umamaheswaran, R. *et al.*, Biomarker signatures in Triassic coprolites. *Palaios*, 2019, **34**, 458–467.
61. Dutta, S. *et al.*, Chemical evidence of preserved collagen in 54-million-year-old fish vertebrae. *Palaeontology*, 2020, **63**, 195–202.
62. Dhiman, H. *et al.*, Discovery of proteinaceous moieties in Late Cretaceous dinosaur eggshells. *Palaeontology*, 2021, **64**(5), 585–595.
63. Dutta, S. *et al.*, Remarkable preservation of terpenoids and record of volatile signalling in plant-animal interactions from Miocene amber. *Sci. Rep.*, 2017, **7**, 1–6.
64. Bhattacharya, S. *et al.*, Amber embalms essential oils: a rare preservation of monoterpenoids in fossil resins from eastern Himalaya. *Palaios*, 2018, **33**, 218–227.
65. Zheng, D. *et al.*, A Late Cretaceous amber biota from central Myanmar. *Nature Commun.*, 2018, **9**, 3170.
66. Paul, S. *et al.*, Preservation of monoterpenoids in Oligocene resin: insights into the evolution of chemical defence mechanism of plants in deep-time. *Int. J. Coal Geol.*, 2020, **217**, 103326.
67. Singh, S. *et al.*, A holistic approach on the gold metallogeny of the Singhbhum crustal province: implications from tectono-metamorphic events during the Archean–Proterozoic regime. *Precambr. Res.*, 2021, **365**, 106376.
68. Kaur, H. *et al.*, Marine microbe as nano-factories for copper biomineralization. *Biotechnol. Bioprocess Eng.*, 2015, **20**, 51–57.
69. Sarkar, S. *et al.*, Spatial heterogeneity in lipid biomarker distributions in the catchment and sediments of a crater lake in central India. *Org. Geochem.*, 2014, **66**, 125–136.
70. Basu, S. *et al.*, Response of grassland ecosystem to monsoonal precipitation variability during the Mid–Late Holocene: inferences based on molecular isotopic records from Banni grassland, western India. *PLoS ONE*, 2019, **14**, 1–24.
71. Ghosh, S. *et al.*, Early Holocene Indian summer monsoon and its impact on vegetation in the Central Himalaya: insight from δD and δ^{13}C values of leaf wax lipid. *Holocene*, 2020, **30**, 1063–1074.
72. Ankit, Y. *et al.*, Long-term natural and anthropogenic forcing on aquatic system – evidence based on biogeochemical and pollen proxies from lake sediments in Kashmir Himalaya, India. *Appl. Geochem.*, 2021, **131**, 105046.
73. Ankit, Y. *et al.*, Apportioning sedimentary organic matter sources and its degradation state: inferences based on aliphatic hydrocarbons, amino acids and δ^{15}N. *Environ. Res.*, 2022, **205**, 112409.
74. Bhattacharya, S. *et al.*, Vegetation history in a peat succession over the past 8000 years in the ISM-controlled Kedarnath region, Garhwal Himalaya: reconstruction using molecular fossils. *Front. Earth Sci.*, 2021, **9**, 703362.
75. Misra, S. *et al.*, Vegetational responses to monsoon variability during Late Holocene: Inferences based on carbon isotope and pollen record from the sedimentary sequence in Dzukou valley, NE India. *Catena*, 2020, **194**, 104697.
76. Basu, S., Ghosh, S. and Sanyal, P., Spatial heterogeneity in the relationship between precipitation and carbon isotopic discrimination in C3 plants: inferences from a global compilation. *Global Planet. Change*, 2019, **176**, 123–131.
77. Agrawal, S. *et al.*, C4 plant expansion in the Ganga Plain during the last glacial cycle: insights from isotopic composition of vascular plant biomarkers. *Org. Geochem.*, 2014, **67**, 58–71.
78. Jha, D. K., Sanyal, P. and Philippe, A., Multi-proxy evidence of Late Quaternary climate and vegetational history of north-central India: implication for the Paleolithic to Neolithic phases. *Quaternary Sci. Rev.*, 2020, **229**, 106121.
79. Sarangi, V., Kumar, A. and Sanyal, P., Effect of pedogenesis on the stable isotopic composition of calcretes and n-alkanes: implications for palaeoenvironmental reconstruction. *Sedimentology*, 2019, **66**(5), 1560–1579.
80. Castañeda, I. S. and Schouten, S., A review of molecular organic proxies for examining modern and ancient lacustrine environments. *Quaternary Sci. Rev.*, 2011, **30**(21–22), 2851–2891.
81. Basu, S. *et al.*, Lipid distribution in the Lake Ennamangalam, South India: indicators of organic matter sources and paleoclimatic history. *Quaternary Int.*, 2017, **443**, 238–247.
82. Roy, B. *et al.*, Morpho-tectonic control on the distribution of C3–C4 plants in the central Himalayan Siwaliks during Late Plio-Pleistocene. *Earth Planet. Sci. Lett.*, 2020, **535**, 116119.
83. Roy, S., Ghosh, S. and Sanyal, P., Carbon reservoir perturbations induced by Deccan volcanism: stable isotope and biomolecular perspectives from shallow marine environment in eastern India. *Geobiology*, 2022, **20**(1), 22–40.
84. Roy, S. *et al.*, Atmospheric CO_2 estimates based on Gondwanan (Indian) pedogenic carbonates reveal positive linkage with Mesozoic temperature variations. *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, 2021, **582**, 110638.
85. Roy, B. *et al.*, Biomarker and carbon isotopic evidence of marine incursions in the Himalayan foreland basin during its overfilled stage. *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 2020, **556**, 109854.
86. Ghosh, S., Sanyal, P. and Kumar, R., Evolution of C4 plants and controlling factors: insight from *n*-alkane isotopic values of NW Indian Siwalik paleosols. *Org. Geochem.*, 2017, **110**, 110–121.
87. Roy, B. *et al.*, Biomarker and carbon isotopic evidence of marine incursions in the Himalayan foreland basin during its overfilled stage. *Paleoceanogr. Paleoclimatol.*, 2021, **36**(5), e2020PA004083.
88. Roy, B. *et al.*, The carbon isotopic composition of occluded carbon in phytoliths: a comparative study of phytolith extraction methods. *Rev. Palaeobot. Palynol.*, 2020, **281**, 104280.
89. Ghosh, S. *et al.*, Substrate control of C4 plant abundance in the Himalayan foreland: a study based on inter-basinal records from Plio-Pleistocene Siwalik Group sediments. *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, 2018, **511**, 341–351.
90. Basu, S. *et al.*, Variation in monsoonal rainfall sources (Arabian Sea and Bay of Bengal) during the late Quaternary: implications for regional vegetation and fluvial systems. *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, 2018, **491**, 77–91.
91. Basu, S., Ghosh, S. and Sanyal, P., Spatial heterogeneity in the relationship between precipitation and carbon isotopic discrimination in C3 plants: Inferences from a global compilation. *Global Planet. Change*, 2019, **176**, 123–131.
92. Roy, S. *et al.*, Atmospheric CO_2 estimates based on Gondwanan (Indian) pedogenic carbonates reveal positive linkage with Mesozoic temperature variations. *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, 2021, **582**, 110638.
93. Sarangi, V. *et al.*, The disparity in the abundance of C4 plants estimated using the carbon isotopic composition of paleosol components. *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, 2021, **561**, 110068.
94. Sarangi, V. *et al.*, Effect of burning on the distribution pattern and isotopic composition of plant biomolecules: implications for paleoecological studies. *Geochim. Cosmochim. Acta*, 2022, **318**, 305–327.
95. Jha, D. K. *et al.*, The first evidence of controlled use of fire by prehistoric humans during the Middle Paleolithic phase from the Indian subcontinent. *Palaeogeogr. Palaeoclimatol., Palaeoecol.*, 2021, **562**, 110151.
96. Ghosh, A. *et al.*, Culture independent molecular analysis of bacterial communities in the mangrove sediment of Sundarban, India. *Saline Syst.*, 2010, **6**(1), 1–11.
97. Bulbul, M. *et al.*, Characterization of sedimentary organic matter and depositional processes in the Mandovi estuary, western India: an integrated lipid biomarker, sedimentological and stable isotope approach. *Appl. Geochem.*, 2021, **131**, 105041.
98. Ajay, K. *et al.*, Distribution and characteristics of microplastics and phthalate esters from a fresh water lake system in central Himalayas. *Chemosphere*, 2021, **283**, 131132.
99. Neelavannan, K. *et al.*, Microplastics in the high-altitude Himalayas: assessment of microplastic contamination in freshwater lake sediments, Northwest Himalaya (India). *Chemosphere*, 2022*,* **290**, 133354.

100. Laskar, N. and Kumar, U., Plastics and microplastics: a threat to environment. *Environ. Technol. Innov.*, 2019, **14**, 100352.
101. Vaid, M. *et al.*, Microplastics as contaminants in Indian environment: a review. *Environ. Sci. Pollut. Res.*, 2021, **28**, 68025–68052.
102. Shivaji, S. *et al.*, Vertical distribution of bacteria in a lake sediment from Antarctica by culture-independent and culture-dependent approaches. *Res. Microbiol.*, 2011, **162**(2), 191–203.
103. Thomas, S. P. *et al.*, Legacy of a Pleistocene bacterial community: patterns in community dynamics through changing ecosystems. *Microbiol. Res*., 2019, **226**, 65–73.
104. Jain S. *et al.*, Ancient DNA reveals Late Pleistocene existence of ostriches in Indian sub-continent. *PLoS ONE*, 2017, **12**(3), e0164823.
105. Parvathi, A. *et al.*, Dominance of *Wolbachia* sp. in the deep-sea sediment bacterial metataxonomic sequencing analysis in the Bay of Bengal, Indian Ocean. *Genomics*, 2020, **112**(1), 1030–1041.
106. Shinde, V. *et al.*, An ancient Harappan genome lacks ancestry from steppe pastoralists or Iranian farmers. *Cell*, 2019, **179**(3), 729–735.
107. Pathak, A. K. *et al.*, The genetic ancestry of modern Indus Valley populations from northwest India. *Am. J. Hum. Genet.*, 2018, **103**(6), 918–929.
108. Harney, É. *et al.*, Ancient DNA from the skeletons of Roopkund Lake reveals Mediterranean migrants in India. *Nature Commun.*, 2019, **10**(1), 1–10.
109. Pörtner, H.-O. (eds) *et al.*, In Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press (in press).
110. Cavicchioli *et al.*, Scientists' warning to humanity: microorganisms and climate change. *Nature Rev. Microbiol.*, 2019, **17**(9), 569–586.
111. Lewin, H. A. *et al.*, Earth BioGenome project: sequencing life for the future of life. *Proc. Natl. Acad. Sci. USA*, 2018, **115**(17), 4325–4333.
112. Gilbert, J. A., Jansson, J. K. and Knight, R., The Earth Microbiome Project: successes and aspirations. *BMC Biol.*, 2014, **12**(1), 1–4.
113. Dabney, J., Meyer, M. and Pääbo, S., Ancient DNA damage. *Cold Spring Harbor Perspect. Biol.*, 2013, **5**(7), a012567.
114. Welker, F. *et al.*, Enamel proteome shows that *Gigantopithecus* was an early diverging pongine. *Nature*, 2019, **576**(7786), 262–265.

ACKNOWLEDGEMENTS. This article arose from a suggestion from the Geoscience Programme Advisory and Monitoring Committee, MoES, Government of India, and multiple discussions thereafter with experts in the country. We thank the anonymous reviewer for insights that helped improve the manuscript.

Received 15 August 2022; re-revised accepted 4 November 2022

doi: 10.18520/cs/v124/i1/26-37