The sustainability of habitability on terrestrial planets: Insights, questions, and needed measurements from Mars for understanding the evolution of Earth-like worlds


Published in:
Journal of Geophysical Research: Earth Surface

DOI:
10.1002/2016JE005134

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
The sustainability of habitability on terrestrial planets: Insights, questions, and needed measurements from Mars for understanding the evolution of Earth-like worlds


1Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California, USA, 2Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA, 3Department of Space Studies, Southwest Research Institute, Boulder, Colorado, USA, 4Department of Earth and Space Sciences, University of Washington, Seattle, Washington, USA, 5School of Earth and Space Exploration, Arizona State University, Tempe, Arizona, USA, 6NASA Marshall Space Flight Center, Huntsville, Alabama, USA, 7NASA Sagan Fellow, 8Department of Physics and Astronomy, Northern Arizona University, Flagstaff, Arizona, USA, 9Department of Astronomy and Carl Sagan Institute, Cornell University, Ithaca, New York, USA, 10Department of Earth and Atmospheric and Planetary Sciences, University of Alberta, Edmonton, Alberta, Canada, 11Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, Indiana, USA, 12Laboratory for Atmospheric and Space Physics, University of Colorado Boulder, Boulder, Colorado, USA, 13The Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland, USA, 14Department of Geosciences, Penn State University, University Park, Pennsylvania, USA, 15Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark, 16Department of Geophysical Sciences, University of Chicago, Chicago, Illinois, USA, 17Solar System Exploration Division, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA, 18Laboratoire Planétologie et Géodynamique de Nantes, UMR6112, CNRS and Université de Nantes, Nantes Cedex 3, France, 19NASA Johnson Space Center, Houston, Texas, USA, 20Department of Earth, Environmental and Planetary Science, Brown University, Providence, Rhode Island, USA, 21Laboratoire de Géologie de Lyon: Terre, Planètes, Environnement, Université Lyon 1/Ecole Normale Supérieure de Lyon/CNRS, Villeurbanne, France, 22Geology Department, Physics and Astronomy Department, Western Washington University, Bellingham, Washington, USA, 23Department of Earth and Planetary Sciences, University of California, Davis, California, USA, 24IRAP, Université de Toulouse, CNRS, UPS, Toulouse, France, 25Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo, Japan, 26Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, 27Centre for Earth Evolution and Dynamics, Department of Geosciences, University of Oslo, Oslo, Norway, 28Harvard Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA, 29Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts, USA, 30School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA, 31Planetary Science Institute, Tucson, Arizona, USA, 32Space Science Division, NASA Ames Research Centre, Moffett Field, California, USA

Abstract
What allows a planet to be both within a potentially habitable zone and sustain habitability over long geologic time? With the advent of exoplanetary astronomy and the ongoing discovery of terrestrial-type planets around other stars, our own solar system becomes a key testing ground for ideas about what factors control planetary evolution. Mars provides the solar system’s longest record of the interplay of the physical and chemical processes relevant to habitability on an accessible rocky planet with an atmosphere and hydrosphere. Here we review current understanding and update the timeline of key processes in early Mars history. We then draw on knowledge of exoplanets and the other solar system terrestrial planets to identify six broad questions of high importance to the development and sustaining of habitability (unprioritized): (1) Is small planetary size fatal? (2) How do magnetic fields influence atmospheric evolution? (3) To what extent does starting composition dictate subsequent evolution, including redox processes and the availability of water and organics? (4) Does early impact bombardment have a net deleterious or beneficial influence? (5) How do planetary climates respond to stellar evolution, e.g., sustaining early liquid water in spite of a faint young Sun? (6) How important are the timescales of climate forcing and their dynamical drivers? Finally, we suggest crucial types of Mars measurements (unprioritized) to address these questions: (1) in situ petrology at
multiple units/sites; (2) continued quantification of volatile reservoirs and new isotopic measurements of H, C, N, O, S, Cl, and noble gases in rocks that sample multiple stratigraphic sections; (3) radiometric age dating of units in stratigraphic sections and from key volcanic and impact units; (4) higher-resolution measurements of heat flux, subsurface structure, and magnetic field anomalies coupled with absolute age dating. Understanding the evolution of early Mars will feed forward to understanding the factors driving the divergent evolutionary paths of the Earth, Venus, and thousands of small rocky extrasolar planets yet to be discovered.

1. Introduction

The advent of exoplanetary astronomy has led to the ongoing detection and characterization of thousands of planetary systems orbiting other stars. Initial data reveal that extrasolar system architecture—planetary size, spatial distribution, and orbital parameters—is often profoundly different from our own solar system [e.g., Malhotra, 2015]. Nonetheless, dozens of planets are known as of mid-2016 whose parameters suggest they might be “Earth-like”: rocky bodies, possibly with gas envelopes, and including some located in a zone where liquid water might be supported at or near the planet’s surface. The concept of a circumstellar “habitable zone” is governed by the last criterion, but planet size, mass, composition, and other dynamical and historical factors must surely play additional roles, as we describe below. On Earth, the limits of life appear largely set by the availability of water with microbial metabolisms and repair mechanisms responding to the challenges of temperature, aridity, radioactivity, and nutrient or energy paucity via a plethora of adaptations. Though subject to debate, the availability of liquid water is the definition of a habitable zone that we will use here.

Statistics of planetary radius, density, atmospheric composition, orbital parameters, and perhaps even atmospheric and surface composition will become known for large numbers of Earth-like exoplanets in the coming few decades. Planets similar in size and mass to our solar system’s planets have already been discovered, and smaller planets will likely be discovered next as the sensitivity of telescopes improves (Figure 1). However, the framework for interpreting their habitability over geologic time, which is crucial for the evolution of complex life, still requires more insights on the factors that control planetary evolution.

Critically, examination of our own solar system demonstrates that a statically defined habitable zone, temporally unchanging, cannot alone explain how likely a planet is to host or have hosted life. To explain why some planets sustain habitable environments over billions of years and some do not requires consideration of planetary and stellar evolution. Key factors include how planets were assembled; the relative role of initial composition versus size in setting the evolutionary pathways of a planet’s magnetic, magmatic, tectonic, and atmospheric evolution; and the stochastic role of impacts or dynamical parameters (Table 1).

On a timescale of decades to centuries, the avenues available to probe the evolution of habitability through time lie in our solar system, making understanding the early history of Earth, Mars, and Venus essential for the evaluation of the potential habitability of terrestrial exoplanets. The work done comparing our solar system’s rocky planets provides the framework for interpreting the hundreds of similar, potentially habitable, worlds predicted to be found around Sun-like stars and smaller main sequence dwarfs.

Here we discuss the history of Mars through the lens of understanding terrestrial planetary evolution and habitability. Many of the questions raised herein are equally applicable to Venus evolution (for a summary of those crucial required measurements, see Venus Exploration Analysis Group [2014] and compare to Mars Exploration Program Analysis Group (MEPAG) [2015]). Likewise, key questions still remain about the initiation and habitability of early Earth’s oceans, paleosurface temperature, and the drivers of the rise of atmospheric oxygen (see for review Catling and Kasting [2017]). We focus on Mars because within our solar system, it exclusively preserves the geologic record required to understand the early evolution of habitability. Unlike Earth and Venus, large swaths of crust from Mars’ first billion years are preserved with stratigraphic context at the surface or near surface, accessible for exploration. In contrast to the Moon, Mercury or icy satellites, the early crustal record of Mars records the evolution of a body with a substantial atmosphere and rocky surface with liquid water. Moreover, Mars’ geologic record is weighted toward the first 1 Gyr of planetary evolution, a period that is disproportionately important for the long-term evolution of planetary habitability because rapid changes in mantle temperature and volatile inventory occur early on.
The Mars exploration program of the last two decades has generated seminal discoveries, including insights into geologic and atmospheric processes supporting ancient habitable environments with liquid water as well as modern day loss and cycling of volatiles. Mars supported multiple aqueous, habitable environments during its first billion years. The main exploration question has transitioned from “Was Mars a habitable world?” to (a) “Was/is Mars inhabited?” and (b) “Why did major planetary-scale transitions in environmental conditions and habitability occur?” With its cold, dry, irradiated surface today, few would argue that Mars is anything but at the edge of habitability, though subsurface refugia for life may still exist [e.g., Cockell, 2014].

Recent choices for Mars exploration have been strongly influenced by the search for biosignatures to address question (a) [e.g., Farmer and Des Marais, 1999; ExoMars Project, 2014; Grotzinger et al., 2012; Mustard et al., 2013]. Certainly, outstanding questions remain about conditions most conducive to preserving any biosignatures and about processes that destroy organics. Here we articulate the set of crucial Mars-related questions and investigations about early evolution, question (b), which has broader relevance across our solar system and to exoplanetary systems.

Figure 1. Solar irradiance as a function of planetary radius for a subset of confirmed and candidate exoplanets. As telescope and processing technologies are refined, smaller planetary candidates are being reported for assessment that are similar to our solar system’s rocky planets. The orange line indicates planetary equilibrium temperatures where liquid water is permitted, assuming a planetary albedo of 0.25 and not considering atmospheric radiative forcing or brines.

Table 1. Key Primary and Derived Parameters for the Evolution of Habitable Terrestrial Planets Along With Resultant Effects

<table>
<thead>
<tr>
<th>Primary Parameter</th>
<th>Exoplanet Measurable?</th>
<th>Derived Parameters</th>
<th>Resultant Effectsa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Distance from star</td>
<td>Yes</td>
<td>Solar irradiation versus wavelength; solar wind particle flux and variability; solar wind particle events; irradiation (EUV, UV, visible, and IR)</td>
<td>Equilibrium surface temperature</td>
</tr>
<tr>
<td>2. Type of star and its flux</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Planetary Size</td>
<td>Yes</td>
<td>Existence/nature of differentiation (magma ocean, overturn) (3, 4, 5); nature and duration of mantle convection (3, 4); nature and duration of core convection (3, 4, 5); existence/duration of plate tectonics (3, 4); volatile flux to the atmosphere (3, 4); core size and composition (3, 4); mantle and crust composition and FeO2 (3, 4); planetary bulk density (4); radiogenic heat flux (4, 5)</td>
<td>Atmospheric loss rate (1–3, 5, 6); heat flux (3–6); magnetic field, including the record of remanent magnetization (3, 4, 5); igneous rock composition and type (3, 4); planetary topography (3, 4, 6); atmospheric composition (1–6); density and density distribution (3, 4, 5)</td>
</tr>
<tr>
<td>4. Starting composition</td>
<td>Partial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Timing of accretion (differences within solar system would lead to differences in parameters 1, 4, 6 by planet)</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Impact bombardment flux and timing</td>
<td>No</td>
<td>Composition of the mantle and crust; impact heating</td>
<td>Melts, breccias, impact-induced decompression lavas</td>
</tr>
<tr>
<td>7. Spin and orbital Parameters</td>
<td>Yes</td>
<td>Time cycle of insolation variability; latitudinal insolation distribution</td>
<td>Variation in surface temperature distribution, climate, and geological processes</td>
</tr>
</tbody>
</table>

aCollectively, the resultant effects dictate the climate (surface temperature, atmospheric circulation, and nature of the hydrologic cycle) and thus surface and near-subsurface environmental conditions, recorded in the geologic record (secondary minerals, physical geology). Resultant effects are italicized when they are determined by multiple primary parameters; this is the case for most, thus necessitating detailed study to disentangle the relative importance of each primary parameter.
Below the current state of knowledge of Mars history is briefly summarized (section 2), key questions about drivers of terrestrial planet evolution are identified (section 3), and, finally, a list of key Mars measurements is provided (section 4), along with a look to future approaches for acquisition of measurements from Mars and from exoplanets (section 5). This paper is intended to be accessible to a diverse audience, including Mars-focused scientists, exoplanet-focused scientists, instrumentalists and exploration program architects. The reader is encouraged to use this outline to focus their attention on sections of greatest interest.

2. The Current View of Mars Evolution and Habitability

We summarize the state of knowledge of Mars and its evolution so as to provide sufficient context for the key questions and measurements. The reader is directed to several other recent, comprehensive reviews on geologic history [Fassett and Head, 2011], mineralogy [Ehlmann and Edwards, 2014; McSween, 2015], sedimentary geology [Grotzinger and Milliken, 2012], magmatism [Grott et al., 2013], Mars tectonics [Golombek and Phillips, 2010], Mars atmospheric evolution [Lammer et al., 2013], and Mars climate [Wordsworth, 2016] that extend this brief review.

2.1. Accretion, Crustal Solidification, and Lack of Mantle Mixing

Timescales of accretion of the solar system’s terrestrial planets from planetary embryos are typically thought to be tens of millions of years in the presence of a large perturbing body (Jupiter), in its absence, at least an order of magnitude more [Morbidelli et al., 2012b]. Indeed, the Moon-forming impact event on Earth—essentially the last collision of the growing Earth with a roughly Mars-sized embryo—is dated to be ~100 Myr after solar system formation [Touboul et al., 2007; Halliday, 2008]. However, this long timescale arises from a need to develop orbit crossing of large embryos. Mars was likely different; it may have been an embryo itself rather than the accumulation of embryos. Indeed, multiple isotopic systems in Martian meteorites indicate that Mars accreted rapidly and at an early stage differentiated into atmosphere, crust, mantle, and core. Mars lacks any volumetrically significant additions ~20 Myr after solar system formation [e.g., Halliday et al., 2011]. Dauphas and Pourmand [2011] from observations based on the Hf-W isotope system proposed that Mars is a planetary embryo with 90–100% of accretion complete by 5 Myr. It is possible that Mars was largely complete while the nebula was still present (unlike Earth and Venus) and thus may have had a hydrogen-rich atmosphere at that time, but this would have been small in mass because Mars is such low mass compared to Earth and Venus.

Mars is less dense than Venus, Earth, or Mercury, implying more volatiles or less iron, or both during accretion. Certainly, compared to Earth, a larger proportion of Mars’ iron resides in the mantle and crust, not core. Mars’ core hosts only ~20% of its mass, while the value is ~30% on Earth. Relatively low abundances of siderophile (iron-loving) elements with chalcophile (sulfur-loving) affinities are consistent with a homogeneous accretion scenario for Mars during which sulfur reacted with metallic iron forming a sulfur-rich FeNi phase for the core [e.g., Treiman et al., 1987]. This likely results from formation under more oxidative conditions, consistent with volatile richness, which may be related to Mars’ position in a colder part of the protoplanetary disk.

The Martian mantle was not well mixed after core-mantle (or metal-silicate) differentiation [Brandon et al., 2000; Foley et al., 2005; Debaille et al., 2007]. Isotopic data from the < 1 Ga shergottite Martian meteorites suggests the end of crystallization of any magma ocean by ~20–60 Myr to preserve observed isotopic anomalies [Kleine et al., 2004; Foley et al., 2005; Borg et al., 2016] and that the mantle reservoirs from which these meteorites were sourced had likely segregated by ~60 Myr or earlier [Marty and Marti, 2002]. On the other hand, nakhlite meteorite geochemistry suggests a multistep solidification process, culminating in a possible turnover of magma ocean crystallization products as late as ~100 Myr [Debaille et al., 2009]. Either scenario is consistent with modeled ~50 Myr timescales for generating a magma ocean through heating by short-lived radioactive isotopes, adiabatic melting, and solid-state cumulate overturn of a density-unstable crust [Elkins-Tanton et al., 2005]. Short timescales of mantle overturn and formation of primary crust are further supported by zircon ages and bulk rock Sm-Nd model ages in the Martian regolith breccia Northwest Africa 7034/7533 [Humayun et al., 2013; Nyquist et al., 2016]. The thermochemical evolution of the Martian mantle, involving convection beneath a stagnant lithosphere [e.g., Ogawa and Yanagisawa, 2011, 2012], must have been insufficiently vigorous to rehomogenize early-formed geochemical reservoirs. The same is true whether the Martian mantle evolved from a fractionated magma ocean or in an alternative (i.e., magma ocean-absent) scenario in which the post-core formation silicate mantle stratified into depleted and enriched reservoirs by melt extraction and depletion of a convecting upper mantle [e.g., Plesa and Breuer, 2014].
Collectively, the data imply that Mars formed early, segregated early, and did not experience any global melting events that significantly remixed mantle reservoirs or crustal components later than 100 Myr.

**2.2. The Impact Record and Its Framework for Geologic Time**

Moving forward several hundred million years, a palimpsest of basins and impact craters on Mars records the later stages of the assembly of the solar system (Figure 2). Fewer than ~10 basins larger than $D > 500$ km are well-preserved, though Mars might host 20 or more buried ancient basins with $D > 1000$ km if “quasi-circular depressions” and “circular thin areas” observed in topographic and crustal thickness data sets result from impacts [Frey, 2008].

Mapping and stratigraphic analysis of Mars’ global geology based on Viking imager data divides the geologic history into three periods (Noachian, Hesperian, and Amazonian) with eight epoch subdivisions [Tanaka, 1986], based on referent geologic units with defined cumulative crater counts (as a function of diameter and count area). This allows relative ages of geologic units across the planet to be compared. The rim rocks of the Hellas basin pin the start of the Noachian. Older, pre-Noachian rocks underlie these surfaces and are exposed by scarps.

Mars has a wide range of crater densities, implying that some of its surfaces were emplaced recently, while others are ancient. The transfer of relative into absolute ages requires knowledge of a crater production function and a chronology model, currently extrapolated by use of scaling factors from lunar crater density and lunar crater ages from Apollo samples. Werner and Tanaka [2011] summarize two such chronologies (detailed in Hartmann and Neukum [2001]) using different methods (Figure 3). Absolute age assignment uncertainties result from the combination of differing crater production functions and chronology models. They vary by ~10% (for surfaces older than about 3 Ga) and up to a factor of 2 (for surfaces younger than 2.5 Ga), depending on what age range and over what crater diameter range the measurement is performed [Hartmann and Neukum, 2001]. However, up to a billion years of additional uncertainty, arising from recent reassessments of the time-dependent lunar cratering rate, has been suggested for the period around 3–3.5 Ga [Robbins et al., 2013; Robbins, 2014].

Even though the absolute timing and duration of many processes are quite uncertain, that Mars is as heavily cratered as it is testifies to the fact that more than half of its surface dates from at least >3.5 Ga [Tanaka et al., 2014], robust even in the face of the uncertainties above. Thus, its rocks record the response of a terrestrial planet to large-scale impact bombardment and the influence of such bombardment on habitability. That Mars’ impact basins have been heavily modified by later volcanic, erosive, and sedimentary processes testifies to active geology well after heavy bombardment.

**2.3. Early Magnetic Field**

The presence or absence of a magnetic field indicates the internal state of a planet and also influences its interaction with stellar processes, which in turn influences, e.g., atmospheric loss, surface radiation. Mars
has no global magnetic field today but does have intense, localized magnetic anomalies in portions of the Noachian southern highland terrains, stronger than typical crustal magnetic anomalies on Earth [Acuña et al., 1999]. This points to the past existence of a dynamo. The field strengths were probably comparable to Earth’s field but are difficult to estimate because of Mars’ different mineralogy (magnetic carrier phases).

A lack of crustal anomalies in materials disrupted by Mars’ four youngest impact basins—Argyre, Isidis, Utopia, and Hellas—has been interpreted to indicate the cessation of a dynamo by the time of these impacts. That is, the iron-bearing phases in impact melt volumes did not become magnetized upon solidifying, which suggests no contemporaneous global magnetic field. The oldest of these four basins, Hellas, is estimated by crater counts to be 3.9 to 4.1 Ga [Werner, 2008]. Furthermore, based on comparison of these four basins to

Figure 3. A timeline of Mars processes and key stratigraphies utilizing (a) the Hartmann [Hartmann, 2005] and (b) the Neukum [Hartmann and Neukum, 2001; Ivanov, 2001] chronology models, as detailed in Michael [2013]. Solar luminosity [Bahcall et al., 2001]; the timing of the magnetic field (see text for discussion); ages of select Martian impact basins [Werner, 2008]; schematic volcanic activity [e.g., Greeley and Schneid, 1991; Grott et al., 2011]; and atmospheric pressure evolution [Hu et al., 2015] model with older pressure profiles interpolated to loss estimates from noble gas and noble gas isotope ratios; last valley network activity [Fassett and Head, 2008]; outflow channel activity [Tanaka, 1997]; meteorite mantle source region(s) [Borg et al., 2016]; the age of materials in the Gale crater catchment sampled by Curiosity [Farley et al., 2013b], ALH84001 crystallization [Lapen et al., 2010], and carbonate formation [Borg et al., 1999]; NWA7034/7533 zircon crystallization [Humayun et al., 2013] are provided. Age units within stratigraphic sections, exposed at past and candidate landing sites, are also provided: Gusev (crater formation [Werner, 2008] and cap lavas [Greeley et al., 2005]); Meridiani [Arvidson et al., 2006, and references therein]; Gale (crater formation) [Thomson et al., 2011]; NE Syrtis/Nili Fossae (Isidis impact and Syrtis lava cap) [Ehlmann and Mustard, 2012]; Jezero (cap rock) [Goudge et al., 2015, and references therein]; Mawrth Vallis (cap rock) [Loizeau et al., 2012]; Holden (crater age) [Grant and Wilson, 2011; Mangold et al., 2012]; Eberswalde (Holden ejecta overlap by delta [Mangold et al., 2012] and Melas [Quentin et al., 2005; Williams and Weitz, 2014]). When cross-cutting relationships with datable units are not clear, this is indicated by “?.” Cross-hatched boxes indicate mega-breccia (with >10 m of stratigraphy) allowing access to earlier time periods.
other impact basins that are magnetized and of similar age (as defined by crater counts), the dynamo may have declined quite abruptly on timescales of < 100 kyr, though time constraints from impact basins are quite coarse [Lillis et al., 2008]. A rapid decline may be consistent with theory and modeling, depending on the nature of core evolution [e.g., Kuang et al., 2008; Breuer et al., 2015]. Carbonate-associated magnetite in ALH 84001 has remnant magnetization inferred to result from a field at least as strong as Earth’s [Weiss et al., 2008]. Its rock formation age of 4.1 Ga [Lapen et al., 2010] and carbonate formation age of 3.9–4.0 Ga [Borg et al., 1999] indicate similar magnetic field timing constraints as the impact basins. Isolated magnetic anomalies in younger Hadriaca Mons [Lillis et al., 2006] and Apollinaris Mons [Langlais and Purucker, 2007] might either suggest a second dynamo epoch at 3.7–3.8 Ga or merely intrusive volcanic rock magnetized in the presence of the field that predated the surface ages recorded by crater counts. Thus, multiple lines of geological evidence support the existence of a strong Martian dynamo and its possible decline by 3.9–4.1 Ga. Nevertheless, until paleomagnetic studies are conducted on rock samples with precise and accurate ages, the timing of the dynamo activity and the reasons for the observed variation in intensities of magnetization of the ancient crust will remain uncertain. The timing constraint on decline is pre-Hellas but post-ALH 84001 carbonate (Figure 3). The implications for atmospheric loss remain to be established (see section 2.6).

2.4. Topography, Volcanism, Tectonism, and Crustal Composition

The primary, global topographic feature on Mars is the north-south hemispheric dichotomy (Figure 2), a >4 km vertical elevation change from southern highlands to northern lowlands over a lateral distance of only a few hundreds of kilometers [Zuber et al., 2000]. Though an origin through either primary crust formation during magma ocean overturn [Elkins-Tanton et al., 2005] or later degree-one mantle convection [Zhong and Zuber, 2001; Roberts and Zhong, 2006] has been considered, an impact origin is currently the prevailing
The modern topography of Mars is testament to a convective pattern within the Martian mantle that may have been stable for most of its existence and that may have prevented mantle flow from destroying mantle isotopic heterogeneities [Kiefer, 2003]. This system has contributed to Mars’ style of volcanism with several large constructs, mons, each 15–20 km tall, ~10^{15} m^3 (twice as large as the Big Island of Hawaii from ocean floor to summit), mostly emplaced within the regional topographic highs at Tharsis and Elysium [Plescia, 2004]. For example, Tharsis rises 10 km above the datum and is surrounded by radial extensional grabens and rifts and concentric compressional wrinkle ridges that together deform the entire western hemisphere and northern plains [Phillips et al., 2001]. The large topographic bulge from emplaced Tharsis lavas may have induced true polar wander in the late Noachian/early Hesperian [e.g., Bouley et al., 2016], though large-scale tectonic structures expected by the stresses induced by such a reorientation have not been found. The emplacement of Tharsis lavas almost certainly contributed to graben formation from normal faulting, including the Tharsis-radial graben swarms [Wilson and Head, 2002; Banerdt and Golombek, 2000; Mége and Masson, 1996] as well as the Valles Marineris canyon system [e.g., Andrews-Hanna, 2012]. The Tharsis province has been growing since the early Noachian [Anderson et al., 2001] and its volcanoes are still intermittently active in the current epoch [Werner, 2009]. At Olympus Mons and Elysium, the most recent lava flows formed ~10–30 Ma [Hartmann and Neukum, 2001]. Many of the shergottite meteorites may source from one of the recent flows on the mons as they have radiometric ages as recent as 165 Ma [e.g., Nyquist et al., 2001]. Plains-style volcanism—effusive, covering large areas but not building substantial mons—has also occurred repeatedly with voluminous Hesperian activity that filled the northern plains [Head et al., 2002] and formed large provinces [e.g., Syrtis Major; Hesperia Planum, and Malea Planum] [Werner, 2009] and with rarer occurrence in the Amazonian.

The vast majority of the crust of Mars is basaltic to basaltic andesite in composition, though distinct modal mineralogies define discrete regions [Rogers and Christensen, 2007; McSween et al., 2009]. At the Gusev and Gale crater landing sites, basaltic lavas and possible pyroclastic deposits of alkaline composition are detected [McSween et al., 2006a, 2006b; Stolper et al., 2013; Sautter et al., 2014, 2015]. Counterparts to these alkali-rich rocks have also been reported within Martian meteorites as igneous clasts in NWA 7034 [Agee et al., 2013; Santos et al., 2015] and as alkali-rich melt inclusions in the chassignite meteorites [Nekvasil et al., 2007, 2009]. These rocks are thought to result from partial melting and fractional crystallization of an oxidized, hydrous mantle source. Basaltic counterparts among Martian meteorites are much more common than the alkali-rich varieties and are represented primarily by the 160–480 Ma shergottite meteorites, which sample Martian basaltic lavas. The shergottites are reported to be the products of partial melting (and subsequent fractional crystallization) of the Martian mantle under fairly reducing conditions (iron-wustite (IW) to fayalite-magnetite-quartz (FMQ)) [e.g., Wadhwa, 2001; Herd et al., 2002] and water-undersaturated conditions [e.g., McCubbin et al., 2016, and references therein]. The nakhlites, chassignites, and ALH84001 meteorites are cumulate rocks (clinopyroxenites, dunites, and an orthopyroxenite, respectively) that likely formed from fractional crystallization of the Martian mantle under fairly reducing conditions (iron-wustite (IW) to fayalite-magnetite-quartz (FMQ)) [e.g., Wadhwa, 2001] and its volcanos are still intermittently active in the current epoch [Werner, 2009]. At Olympus Mons and Elysium, the most recent lava flows formed ~10–30 Ma [Hartmann and Neukum, 2001]. Many of the shergottite meteorites may source from one of the recent flows on the mons as they have radiometric ages as recent as 165 Ma [e.g., Nyquist et al., 2001]. Plains-style volcanism—effusive, covering large areas but not building substantial mons—has also occurred repeatedly with voluminous Hesperian activity that filled the northern plains [Head et al., 2002] and formed large provinces [e.g., Syrtis Major; Hesperia Planum, and Malea Planum] [Werner, 2009] and with rarer occurrence in the Amazonian.

The vast majority of the crust of Mars is basaltic to basaltic andesite in composition, though distinct modal mineralogies define discrete regions [Rogers and Christensen, 2007; McSween et al., 2009]. At the Gusev and Gale crater landing sites, basaltic lavas and possible pyroclastic deposits of alkaline composition are detected [McSween et al., 2006a, 2006b; Stolper et al., 2013; Sautter et al., 2014, 2015]. Counterparts to these alkali-rich rocks have also been reported within Martian meteorites as igneous clasts in NWA 7034 [Agee et al., 2013; Santos et al., 2015] and as alkali-rich melt inclusions in the chassignite meteorites [Nekvasil et al., 2007, 2009]. These rocks are thought to result from partial melting and fractional crystallization of an oxidized, hydrous mantle source. Basaltic counterparts among Martian meteorites are much more common than the alkali-rich varieties and are represented primarily by the 160–480 Ma shergottite meteorites, which sample Martian basaltic lavas. The shergottites are reported to be the products of partial melting (and subsequent fractional crystallization) of the Martian mantle under fairly reducing conditions (iron-wustite (IW) to fayalite-magnetite-quartz (FMQ)) [e.g., Wadhwa, 2001; Herd et al., 2002] and water-undersaturated conditions [e.g., McCubbin et al., 2016, and references therein]. The nakhlites, chassignites, and ALH84001 meteorites are cumulate rocks (clinopyroxenites, dunites, and an orthopyroxenite, respectively) that likely formed from basaltic magmas or lavas [e.g., McSween, 2008], although the chassignites and nakhlites have many geochemical similarities indicating they may be sourced from the same thick lava flow or shallow intrusion [McCubbin et al., 2013]. Additional cumulates have been excavated by impact craters [Skok et al., 2012]. Multiple studies have shown that low-calcium pyroxene-bearing lavas are common in Noachian terrains, late-Noachian and later lavas are dominated by high-calcium pyroxene and olivine [e.g., Mustard et al., 2005; Flahaut et al., 2012], and the geochemistry of Martian igneous rocks changes with time [Baratoux et al., 2011; Filiberto and Dasgupta, 2015]. This evolution in composition has been hypothesized to indicate cooling of the Martian mantle and consequent changes in the degree of partial melting [Baratoux et al., 2011, 2013]. Also, the ancient materials (Gusev, Gale, and NWA 7034) seem to be more hydrous and oxidizing while the younger shergottites are reduced and drier. This is either a sign of evolution of the mantle or sampling of heterogeneous reservoirs [e.g., Balta and McSween, 2013; Schmidt et al., 2013; Tuff et al., 2013].

There are hints of additional igneous diversity in the Martian rocks. For example, ancient, high magnesium olivine-enriched units are circumferential to the Isidis and Argyre impact basins, possibly ultramafic, and point to excavation of the mantle or post-impact lavas with very high degrees of partial melting [Koeppen and Hamilton, 2008; Ody et al., 2013]. More silicic, evolved rocks have also been reported in small exposures from orbit [Christensen et al., 2005; Wray et al., 2013; Carter and Poulet, 2013] and in rocks with large feldspar...
phenocrysts in conglomerates at Gale crater [Sautter et al., 2014]. The regolith breccia NWA7034 (paired with NWA 7533 and assembled ~1.7 Ga) includes more evolved trachyandesite and Fe-Ti-P-rich lithologies, not otherwise represented among the Mars meteorites [Santos et al., 2015]. More recently, tridymite, possibly a product of high-temperature, low-pressure silicic volcanism, has been identified at Gale crater [Morris et al., 2016]. Thus, while basaltic volcanism has dominated for much of the last 3.5 Ga, a more complex set of igneous processes operated earlier on Mars.

2.5. The Record of Water and Environments

Erosion rates were substantially greater during the Noachian than on modern Mars, pointing to enhanced early aeolian and fluvial activity [e.g., Golombek et al., 2014, and references therein]. Dendritic valley networks in equatorial and midlatitudes clearly indicate drainage of surface waters, and the most extensive regionally integrated systems were last active in the late Noachian-to-early Hesperian [Howard et al., 2005; Irwin et al., 2005; Fassett and Head, 2008]. Some of these valleys terminate in open- or closed-basin lakes, hosted in craters [Goudge et al., 2016, and references therein]. A northern ocean has been proposed based on irregular paleoshorelines (see for review Carr and Head [2003]) and fans and deltas located along the dichotomy boundary [Di Achille and Hynek, 2010; DiBiase et al., 2013].

By the late Hesperian/early Amazonian, the type of liquid water-related landforms had changed: dendritic valleys became restricted in extent, though paleolakes and related fluvial landforms were still locally and episodically present at the surface [e.g., Quantin et al., 2005; Mangold et al., 2007, 2008, 2012; Grant and Wilson, 2011; Mangold, 2012; Ehlmann and Buz, 2015]. Large outflow channels debouched into the northern lowlands, apparently the result of catastrophic floods from sources not fully understood but probably related to buried aquifers or rapid melting of surface or ground ice [Tanaka, 1997]. Amazonian landforms still include local observations of fluvial landforms, including poorly dendritic valleys, supraglacial channels, and recent gullies [Malin and Edgett, 2000; Costard et al., 2002; Schon et al., 2009; Fassett et al., 2010; Mangold, 2012; Kite et al., 2013a, 2013b] and intriguingly, slope lineae, possibly water related [McEwen et al., 2011]. A change in the character and decline of surface water-related features through time is apparent from the more patchy distribution and smaller size of fluvial landforms.

The rich mineral record of water-rock interactions also shows broad trends with time [e.g., Bibring et al., 2006]. Iron/magnesium phyllosilicates are widespread in Mars’ oldest rocks, exposed beneath present-day surface units and pointing to globally widespread water-rock alteration in the Noachian and earlier. The setting of this alteration—weathering, aquifer-hosted, hydrothermal or deuteric—is debated [Poulet et al., 2005; Mustard et al., 2008; Meunier et al., 2012]. Much of the mineralogic alteration to form widespread Fe/Mg phyllosilicates appears to be approximately isochemical [Taylor et al., 2010; Ehlmann et al., 2011a; Arvidson et al., 2014; McLennan et al., 2014]. Mineral assemblages in some Noachian locations include phases formed by subsurface alteration in diagenetic, mildly hydrothermal, or subgreenschist metamorphic conditions [Ehlmann et al., 2009, 2011a, 2011b; Carter et al., 2013]. Near-surface volcanic hydrothermal or fumarolic activity to form silica deposits has been found on the surface at Gusev crater and in Hesperian units at Syrtis Major [Spudis et al., 2008; Skok et al., 2010]. A characteristic chemical stratigraphy of Fe/Mg phyllosilicates, overlain by units rich in Al-phyllosilicates, which are in turn capped by nonaltered units formed in the early Hesperian, may have formed from aqueous alteration of less mafic starting compositions, intensive near-surface leaching of basaltic materials from waters that were cumulatively voluminous or acidic, or both (e.g., Murchie et al. [2009], Ehlmann et al. [2011a], and Carter et al. [2015]).

Salts are found in several basins, indicating open system weathering and water chemistry that varied in time and space, inducing precipitation of carbonates, chlorides, and sulfates, including acid indicators jarosite and alunite (e.g., see Ehlmann and Edwards [2014] for review). Timing of many deposits is late Noachian to early Hesperian but may be considerably later for some basins with sulfates and chlorides [Gendrin et al., 2005; Osterloo et al., 2010]. Gale crater has evidence for multiple lake episodes around the late Noachian/early Hesperian as well as longer-lived diagenetic fluid alteration to form clays and sulfates [e.g., Grotzinger et al., 2014, 2015]. Widespread plains of hematite- and sulfate-bearing sediments explored in situ at Meridiani Planum show evidence for an intermittently wet environment in the early Hesperian with alternating playa strata and multiple episodes of groundwater diagenesis [e.g., Grotzinger et al., 2005; McLennan et al., 2005]. Some of the most recent hydrated minerals visible from orbit, which formed in the late
Hesperian/early Amazonian, are associated with the Valles Marineris area near Tharsis [Milliken et al., 2008; Thollot et al., 2012].

Thus, stratigraphic sequences from the first billion years of Mars history comprise a rich and diverse record of water-related environmental conditions on a changing planet. Yet extensive aqueous activity largely ceased by circa 3.0–3.5 Ga. Water-limited, thin-film acid alteration of basaltic rocks continued, producing coatings and rinds on Hesperian basaltic rocks at Gusev crater and possibly Amazonian northern plains sediments [Hurwitz and McLennan, 2007; Horgan and Bell, 2012]. Sands contain trace salts including sulfates, perchlorates, nitrates, and carbonates [Clark and van Hart, 1981; Hecht et al., 2009; Boynton et al., 2009; Leshin et al., 2013], either formed from modern atmospheric processes or eroded and redistributed from older deposits. Evidence for oxidative weathering, which may or may not involve water, is widespread in the dust of Mars [e.g., Morris et al., 2006]. Neither the drivers of spatial and temporal diversity on early Mars nor the reasons for the timing of the decline are understood, but they likely relate to the interplay of impact, volcanic, and atmospheric processes.

### 2.6. Atmospheric Evolution

Mars has a 6 mbar atmosphere today, though all the highly volatile elements (H₂O, CO₂, N₂, and the noble gases) are substantially less abundant on Mars than they are on Earth. Most atmospheric species are strongly mass fractionated with respect to potential sources. This points to substantial atmospheric loss over time.

A prevailing model divides loss of Mars’ atmosphere into two stages [e.g., Bogard et al., 2001; Pepin and Porcelli, 2002, and references therein]. The composition of the primordial atmosphere was set by impact degassing and outgassing. Extreme ultraviolet (EUV) radiation and the solar wind were stronger from the early Sun due to the dynamo amplifying effect of its faster rotation. Absorption of EUV in the upper atmosphere of early Mars should have led to hydrodynamic escape, as hydrogen escape fluxes are thought to have been large enough to exert upward drag forces on heavier atmospheric constituents sufficient to lift them out of the atmosphere. Observed Xe isotope fractionation suggests its loss and loss of most (>99%) of lighter atmospheric constituents. Heavy bombardment by impacts may also have stripped the atmosphere [Melosh and Vickery, 1989] or provided key sources of volatiles [Owen and Bar-Nun, 1995] (see section 3.5.2 for discussion of this point). After the decline of the magnetic field, sputtering loss likely persisted from the Noachian onward [Jakosky et al., 1994; Brain and Jakosky, 1998]. Ongoing degassing of the Mars crust and mantle and, possibly, some delivery of cometary volatiles replenished the supply. The details of early loss (timing, initial reservoir, and process rates) are still a subject of considerable debate as initial inferred Martian Xe ratios, isotopic systematics of the lighter noble gases, and inferred amounts of degassing are not yet fully self-consistent [e.g., Pepin, 2006], and some have recently posited lower early hydrodynamic Xe loss on Earth [Pujol et al., 2011].

Nevertheless, by the late Noachian (~3.8 Ga), hydrodynamic escape had certainly ceased along with loss due to large impacts. Though details of the atmospheric reservoir at that point are uncertain, subsequent removal of volatiles then was likely dominated by nonthermal processes, including pickup ion sputtering, the rate of which depends on the solar wind particle flux [Leblanc and Johnson, 2002], and photochemical processes, in which the rate of loss scales to the solar Lyman continuum flux [Groeller et al., 2014]. This is similar to the loss regime today. Outgassing from the mantle during volcanism added volatiles with the amount and composition of gases released dependent on the fugacity of the mantle, composition of lavas, the rates of intrusive and extrusive volcanism [e.g., Grott et al., 2011], and atmospheric pressure [Gaillard and Scaillet, 2014] (sections 3.1 and 3.3.2). Some sequestration of CO₂ in carbonate-bearing rocks also occurred [Niles et al., 2013; Edwards and Ehlimann, 2015; Wray et al., 2016], though early large amounts of carbonate predicted [Pollack et al., 1987] have not so far been discovered, at least in rocks from the late Noachian onward.

By the late Noachian/early Hesperian, multiple lines of evidence point to Martian atmosphere being relatively thin, i.e., ~<2 bar and likely ~<1 bar [e.g., Kite et al., 2014; Edwards and Ehlimann, 2015; Bristow et al., 2016; Lapotre et al., 2016]. However, D/H measurements, and possibly ^{36}Ar/^{38}Ar measurements, combined with ongoing Mars Atmospheric and Volatile Evolution (MAVEN) measurements of the altitudes of the homopause, exobase, and thermosphere temperatures allow models that suggest that 50–90% of the atmosphere is lost over 4 Ga [Atreya et al., 2013; Mahaffy et al., 2015; Slipski and Jakosky, 2016]. Hu et al. [2015] recently modeled all sources and sinks finding self-consistent scenarios of decline from ~0.2 to 1.8 bar (likely
<1 bar) over the last 3.8 Ga to the present-day value of 6 mbar, indicated by enriched carbon isotopes (Figure 3). Obliquity changes on Mars and the discovery of >5 mbar of CO₂ ice trapped in the southern polar cap [Phillips et al., 2011] along with modeling of atmospheric collapse [Soto et al., 2015] indicate that the modern 6 mbar atmosphere is not a fixed quantity but rather may fluctuate between ~1 mbar and ~30 mbar (see section 3.6).

The influence of atmospheric evolution on climate and the challenges in maintaining liquid water on the surface were recently reviewed by Wordsworth [2016]. In brief, 3-D climate models and several data sets suggest a cold and relatively dry steady state climate from the late Noachian onward with more water ice in higher altitude regions than today. Lack of evidence for glaciation disfavors the long-term existence of a northern ocean, which would have been a water source for an integrated hydrologic system, creating bands of intense precipitation or ice deposits in the southern highlands Wordsworth [2016]. Yet the exact nature of the climate is much debated. Warmer interludes or epochs clearly existed to form valley networks, breach crater lakes, and heavily incise and degrade craters, particularly in the late Noachian and earlier. Whether water that formed valley networks was from rain or snow/ice melt remains unresolved; a key question is how much annual precipitation of either form was available to generate flows? Warmer climates may have been triggered by impacts, volcanism, orbital forcing, or some combination, possibly involving clouds [Toon et al., 2010; Segura et al., 2013; Urata and Toon, 2013; Kite et al., 2013a]. Trace gases available in the atmosphere and proposed to contribute to warming include H₂, H₂O, and, for transient episodes, sulfurous species [Johnson et al., 2009; Ramirez et al., 2013; Halevy and Head, 2014; Wordsworth, 2016], though for sulfurous species, aerosol formation may instead induce cooling [Tian et al., 2010]. Fortunately, the preserved geologic and geochemical record allows these questions about the processes governing atmospheric evolution to be resolved.

3. Questions on Terrestrial Planet Evolution and the Sustainability of Habitability

As observations of terrestrial exoplanets accumulate and characterization of their parameters becomes more complete, these beg the question: what matters and what does not in setting planetary climate and the availability of water, nutrients, and energy sources? The 2014–2023 National Academies Decadal Survey acknowledged this issue with a fundamental question of the “Building New Worlds” theme: “What governed the accretion, supply of water, chemistry, and internal differentiation of the inner planets and the evolution of their atmospheres, and what roles did bombardment by large projectiles play? Important objects for study: Mars, the Moon, Trojans, Venus, asteroids, and comets.” Specifically relevant to the sustaining of Martian habitability through time, the Survey also identified for Mars a key question: “How have the factors and processes that give rise to habitable conditions at planetary and local scales changed over the long term in concert with planetary and stellar evolution?” Here we unpack these cross-cutting questions into high-level measurable and modelable parameters and more specific key questions.

Seven primary parameters, several of which are measurable for exoplanets with telescopic observations, drive planetary evolution; their combined effects on planetary evolution and the sustaining of habitable environments need disentangling (Table 1 and Figure 4). Below we identify six key questions (Table 2) related to these primary parameters, their combined effects, and the evolution of habitability, relevant to understanding the evolution of all terrestrial planets. Significant progress is possible via observation, modeling, and, for many, uniquely accessible by exploration of early Mars’ geologic record (discussed in section 4).

3.1. Is Small Size Fatal?

Size—mass and volume—is certainly one of the easiest parameters to identify for exoplanets. Among terrestrial planets, Mars may be atypically small (Figure 1) or future advances in telescopic technologies may yet discover many Mars-sized exoplanets. Most current models of solar system formation predict the initial generation of a suite of planetary embryos (diameter ≥1000 km). In traditional rocky planet formation models, collisions of these embryos over tens of millions of years generated Earth-mass terrestrial planets. Venus- and Earth-sized bodies are the norm; the small size of Mars, though not impossible to reproduce, is improbable [Raymond et al., 2009]. Thus, Mars formation scenarios that propose a depletion in disk mass around ~1.5 AU have been invoked, due to either different rates of material infall into the Sun as a function of radial distance [Izidoro et al., 2013] or to a so-called “Grand Tack” whereby the orbits of Jupiter and Saturn sweep inward then outward through the solar system, scattering planetesimals and embryos [Walsh et al., 2011]. More recently, models with
rapid accretion from sub-meter scale objects, so-called pebble accretion, are able to reproduce the basic structure of the inner solar system, including a small Mars and low mass asteroid belt [Levison et al., 2015]. Some results from the past decade of exploration, considered together with concepts of habitability, suggest that Mars’ small size caused it to “die out” geologically “too early” for life to have evolved. That is, decreased activity might have facilitated atmospheric loss, making surface habitats too cold or too arid, while a decline in geothermal heat flux might have driven subsurface habitats to freeze out. Such a line of thought belies the fact Mars remains volcanically active today [Hartmann and Neukum, 2001; Werner, 2009], has significant obliquity-driven climate excursions of a scale substantially greater than those of Earth [Jakosky and Carr, 1985; Jakosky et al., 1995; Laskar et al., 2002, 2004; Mustard et al., 2001; Head et al., 2003], and hosts some landforms potentially created by liquid water over the last millions of years [Schon et al., 2009; McEwen et al., 2011]. Additionally, the mode of convection is critical to setting cooling rate. Nevertheless, size is a driver for a number of geophysical processes relevant to habitability for which small size is detrimental (Table 1). Two key questions are found below.

3.1.1. Do Habitable Planets Need to be Well-Stirred?

Mars is sufficiently large that it almost certainly relies on mantle convection to remove heat from the interior. However, there is no unambiguous evidence that Mars had plate tectonics at any epoch. In this regime, called “stagnant lid,” small convecting planets do not cool faster than large convecting planets but operate at a lower internal temperature [Stevenson, 2003]. They may therefore have less volcanic activity, because this depends on upwelling material crossing the mantle solidus at depth; however, the depth of the solidus depends also on mantle composition. What was the volume flux of volcanism on this small planet? Did Mars ever have a form of crustal recycling? All isotopic chronometers reveal very early isotopic heterogeneity in the mantle of Mars that is mostly not preserved on Earth, implying early melting and lack of subsequent large-scale mixing for Mars (section 2.2) [e.g., Halliday et al., 2001]. The isotopic evidence coupled with observed large volcanic constructs leads to the seemingly contradictory conclusion of a convecting mantle that retains primordial heterogeneity. These observations could be reconciled by mantle convection that does not cross compositional stratification in either the vertical (e.g., upper versus lower mantle) or horizontal (e.g., northern lowlands versus southern highlands) directions, perhaps maintained by long-term stable patterns of convection (see section 2.4). There is no unambiguous evidence for the establishment of an Earth-like tectonic cycle (the consequences for habitability are addressed below).
The existence of large volcanos, and indeed lavas formed in the past 100 Myr (see section 2.4), certainly means volcanism has played an ongoing role in releasing volatiles to “refresh” the Martian atmosphere. But what about their eventual fate and possible cycling? Carbonate-silicate weathering cycles, biologically mediated on Earth [e.g., Walker, 1986], have played a role in regulating terrestrial climate. However, Mars is the case where geochemical cycles may have been one way: from source to permanent sink or loss to space. In the absence of unequivocal evidence for plate tectonics, there is not a clear mechanism to cycle hydrogen, carbon, sulfur, nitrogen, and other atmospheric components between the interior and atmosphere. Delamination of the lower crust during early crustal formation [e.g., Morschhauser et al., 2011; Ogawa and Yanagisawa, 2012; Plesa and Breuer, 2014] in the pre-Noachian could have resulted in limited volatile cycling between the interior and surface. However, since the onset of the geological record, the volatile abundances in the interior and atmosphere have likely been controlled by the rates of volcanism, loss to space, and physical weathering in making new surfaces available for chemical reaction. It is unclear, however, that this is inherently fatal to habitability. Mars offers an opportunity to study geochemical reservoirs and weathering without plate tectonics. The establishment of plate tectonics, more specifically the onset of plate subduction, on Earth is imperfectly understood, and its lack may in fact be the “default,” long-term terrestrial planetary scenario (e.g., see also Mercury, Venus, and the Moon). On Venus, outgassing appears to have produced a runaway greenhouse; on Mars, it may be balanced mostly by loss to space.

3.1.2. How Does Size Affect Atmospheric Speciation and Loss?

Mars Express and MAVEN are starting to reveal Mars’ current atmospheric loss rate and its dependence on solar variability [e.g., Jakosky et al., 2015a, 2015b], but the further back in time one goes, the more challenging extrapolation of these measurements becomes because changes in the Sun’s properties suggest hydrodynamic escape and EUV flux were substantially different than today (see sections 2.6 and 3.2).

Many models for planetary atmospheres assume that primary atmospheres are determined mostly by the composition of accreted materials and secondary atmospheres from the composition of gases released during a magma ocean or later volcanic outgassing. However, size may exert a disproportionate influence through gravity, which influences the ease of escape. This in turn dictates atmospheric pressure, which then feeds back to the composition of gases released by volcanism and their susceptibility to further escape. From modeling gas release from terrestrial mid-ocean ridge basalt mantle as an illustrative example, Gaillard and Scalet [2014] suggest that atmospheric pressure itself may be the strongest control on atmospheric composition with multi-bar atmospheres thermodynamically favoring release of N2- and CO2-rich dry volcanic gases, pressures near 1 bar promoting H2O rich gases, and lower pressure promoting sulfur-rich gases. This in turn dictates the level of greenhouse warming and the nature of surface-atmosphere chemical reactions. Mars is just small enough for carbon and oxygen to escape from photolysis of CO in photochemical processes, thus providing an additional potential pathway for its loss relative to Earth and Venus. If the rate of early atmospheric loss was relatively rapid
Secondly, Mars’ record provides the ability to interrogate planetary materials immediately before and after the cessation of the dynamo, looking for signatures of the effects of the existence of a magnetic field on atmospheric species and their loss as well as the preservation of organics/radiation damage in rocks from before and after. The presence of a magnetic field creates complex interactions between the solar wind and the atmosphere, the study of which is an active area of space physics research, including at Mars by...
the ongoing MAVEN mission. The magnetic field is typically thought to prevent sputtering from solar particle events in which charged particles interact with the upper levels of the atmosphere to remove species via sputtering by pickup ions [e.g., Kass and Yung, 1995]. A global dipole magnetic field like the Earth’s causes the solar wind to stand off at a greater distance than it would in the absence of a field. This can protect the atmosphere from stripping by the solar wind [e.g., Hutchins et al., 1997a, 1997b]. But at the same time, it can create pathways for atmospheric loss, where the magnetic field at the cusps can provide direct pathways for ion loss where the field lines are open to the tail (e.g., Earth’s poles) [Strangeway, 2005]. When a magnetic field is present, the magnetosphere can funnel the solar wind into a small auroral zone. The resulting energy flux over the small fraction of the atmosphere could reach 10–100 times larger than in the case with no magnetosphere, thereby promoting the escape of atmospheric gases [e.g., for Earth, Engwall et al., 2009]. The effect is to greatly enhance the escape of heavy species (e.g., O+) from a deep gravitational well like the Earth’s; the impact on light gases is minimal. This enhancement effect is also not believed to be important for Mars, where the atmosphere is much less tightly bound than at Earth [Brain et al., 2013].

On Mars, it is suspected that the turnoff of a global field at approximately 4 Ga allowed stronger interaction between the solar wind and the upper atmosphere and resulted in a turn-on of solar wind stripping of the atmosphere [e.g., Jakosky et al., 2015a, 2015b]. Additionally, absent a magnetic field, a higher flux of charged particles, especially cosmic rays, reaches the surface, an effect which destroys chemical bonds of materials near the surface, may generate oxidants like perchlorates, and may inhibit the preservation of organics and minerals in pristine form in the upper meter [e.g., Pavlov et al., 2012; Kim et al., 2013; Gerakines and Hudson, 2013; Gerakines and Hudson, 2015]. A thinning atmosphere may, however, be a more powerful driver of these effects than loss of a magnetic field.

Because of the completeness of the Martian record with regard to the evolution of the magnetic field, these key aspects—fundamental controls on why some planets sustain magnetic fields and how magnetic field existence and loss affect processes relevant to habitability—may be fully addressed for the Martian case by examination of the rock record.

3.3. Is Accretion Composition Destiny? The Role of Initial Metal:Rock:Volatile Ratios

Telescopic observations show that different stellar environments have different metallicities (ratio of hydrogen to all other elements) and CO ratios, including variability in the distribution of the elements throughout protoplanetary disks. These differences, in turn, dictate the potential materials—ratios of metals, rocky materials, volatiles, and their redox state—available to accrete planets, feeding forward to consequences for igneous and atmospheric evolution. The volatile content and oxygen fugacity (fO2) of the mantle influence the nature and composition of the original crust, phase transitions, core formation, element partitioning, and the rheology of the mantle. This is turn has consequences for the likelihood of tectonics (discussed in section 3.1.1), the composition of the atmosphere, and available geochemical cycles.

3.3.1. Untangling the Controls on Accretion on Volatile Availability

Mars’ likely status as a planetary embryo (see section 2.1) means that it records the accretion and early state of a protoplanet. This may or may not be common for rocky planets with atmospheres (it is not the case for Earth; Venus is unknown). Mars offers an opportunity to characterize the degree of mixing of solar system materials, undisrupted by a relatively late lunar giant impact that mixed embryos. The oxygen isotopic character of Mars is distinctively different from Earth (and Moon), but since we do not know Venus, it is not clear what this tells us about the reservoirs of material used in the accretion of Mars. By systematic sampling of some of Mars’ oldest crust, the possible asteroid belt (meteorite) and cometary building blocks can be assessed and compared to Earth [e.g., Alexander et al., 2012] as well as dynamical models [Lunine et al., 2003]. Competing models for the overall architecture of planet formation [e.g., pebble accretion; Levison et al., 2015], standard models [Chambers, 2004], and the Grand Tack [Gomes et al., 2005] make different predictions for the growth rate and spatial distribution of source material for Mars analog planets, including their volatiles. This insight into early accretion is critical to understand volatile delivery and incorporation into rocky planets. Are the key elements for habitability established on protoplanets or are later accretion processes important/necessary for sustained habitability?

A late oxidized “veneer” component is typically considered necessary to supply volatiles and produce observed, near-chondritic relative abundances of highly siderophile elements on Earth [e.g., Halliday et al., 2001]. Current data suggest any late veneer was accreted to Mars to a lesser extent [Brandon et al., 2012].
The inferred range of 1000°C, the yellow ovals represent the McCubbin et al. [2013] (Figure 6). This relatively low quite different, dominated by CO₂. Based on this observation, we know that the Martian surface becametribution to the total pressure of the directly degassed atmosphere. The present-day Martian atmosphere isPlanetary mantle

tations [e.g., Hirschmann and Withers (2008), Ulmer and Luth (1991), Sharp et al. [2013], and Movrogeneres and O’Neill (1999)]. The blue shaded area represents the habitability region in fO₂, in this case defined as the region allowing for the coexistence of substantial H₂O as well as reduced carbon species. This region extends 1 log unit below the unity point for H₂ and H₂O and 1 log unit above the high-pressure unity point for graphite and C-O-H fluids.

and/or N. Dissolved volatiles in the silicate portion of a planet equilibrate with the metal core. For example, H₂O reacts with metallic Fe⁰ to produce FeO and H₂. This reaction proceeds until the fugacity ratio of H₂ to H₂O has the same oxygen fugacity as that defined by the activity of FeO and Fe⁰ in the silicate mantle and core, respectively [e.g., Hirschmann et al., 2012]. The initial oxygen fugacity (fO₂) of the Martian mantle at the time of core formation is estimated from Martian meteorites to be ~1.25 log units below the iron-wüstite (IW) buffer, which results in an molar H₂/H₂O fugacity ratio between 1 and 10, depending on temperature [Sharp et al., 2013] (Figure 6). This relatively low fO₂ and elevated fH₂ promote the formation of reduced C-H species in silicate melts [Ardia et al., 2013] and destabilize the formation of more oxidized C-O melt species (e.g., CO₃²⁻) [Stanley et al., 2011, 2012]. For Mars, this process seems to have resulted in the production of graphitic carbon in the mantle [Hirschmann and Withers, 2008; Righter et al., 2008] as well as the production of organic macromolecular carbon, including polycyclic aromatic hydrocarbons [Steele et al., 2012]. Nitrogen and sulfur also primarily exist in the reduced form under low fO₂ and elevated fH₂, so S in the Martian mantle likely exists primarily as sulfides and N may exist as nitride phases or as ammonium (NH₃⁺) components in K-rich minerals.

Planetary mantle fO₂ (Figure 6), set by initial composition, differentiation, convection, and impact history, is important because it controls the speciation of volcanic gases and chemical character of the primordial atmosphere. If a magma ocean on Mars degassed a primordial atmosphere, this early atmosphere was likely dominated by H₂ with lesser contributions from H₂O, H₂S, HCl, and CH₄, based on the inferred oxygen fugacity of Mars during differentiation. Species like CO, CO₂, N₂, and SO₂ would have made a very minor contribution to the total pressure of the directly degassed atmosphere. The present-day Martian atmosphere is quite different, dominated by CO₂. Based on this observation, we know that the Martian surface became more oxidized with time. In fact, this increase in oxygen fugacity with time can also be inferred from samples of the Martian crust. The large fO₂ range in the shergottite meteorites from ~IW to ~5 log units above IW, coupled with their geochemistry, indicates that there is an oxidized crust-like component on Mars and a reduced mantle component [e.g., Wadhwa, 2001; Herd et al., 2002; Shearer et al., 2013; Steele et al., 2012].

The inferred range of fO₂ of the ~3.7 Ga Gusev basalts is similarly large but extends to more oxidized conditions [Schmidt et al., 2013; Tuff et al., 2013]. Finally, the most representative sample of Martian crust that we

Figure 6. Estimated ranges of oxygen fugacity among a subset of terrestrial bodies, based on ranges reported in the literature as derived from analyses of samples [Sharp et al., 2013; Papike et al., 2005; Zolotov et al., 2013; McCubbin et al., 2012; Wadhwa, 2001; Herd et al., 2002]. At temperatures of 1000°C, the yellow ovals represent the fO₂ (relative to IW) at which the oxidized and reduced forms of a volatile pair have an activity ratio of unity (from Hirschmann and Withers [2008], Ulmer and Luth [1991], Sharp et al. [2013], and Movrogeneres and O’Neill [1999]). The blue shaded area represents the habitability region in fO₂, in this case defined as the region allowing for the coexistence of substantial H₂O as well as reduced carbon species. This region extends 1 log unit below the unity point for H₂ and H₂O and 1 log unit above the high-pressure unity point for graphite and C-O-H fluids.

Thus, for Mars, its accretion composition may have largely dictated destiny. The essential information to understand the source regions and timing of accreted materials is recorded in quantification of Mars’ volatile reservoirs and the isotopic systematics of Mars mantle components, tapped by volcanism.

3.3.2. What Controls the Planetary-Scale Availability of H₂O and Organics?
A rocky planet’s initial redox state is linked to both the composition of its metallic core and the abundances and oxidation states of its initial inventory of volatiles. The interactions (reaction, cycling, or loss) of redox-sensitive elements, H, C, N, O, S, and Fe in planetary reservoirs have profound implications for thermal, magmatic, and atmospheric evolution by controlling phase speciation. Most core compositions in the inner solar system are predominately Fe-rich with additional components that include Ni, S, C, Si, and possibly H

Planetary mantle fO₂ (Figure 6), set by initial composition, differentiation, convection, and impact history, is important because it controls the speciation of volcanic gases and chemical character of the primordial atmosphere. If a magma ocean on Mars degassed a primordial atmosphere, this early atmosphere was likely dominated by H₂ with lesser contributions from H₂O, H₂S, HCl, and CH₄, based on the inferred oxygen fugacity of Mars during differentiation. Species like CO, CO₂, N₂, and SO₂ would have made a very minor contribution to the total pressure of the directly degassed atmosphere. The present-day Martian atmosphere is quite different, dominated by CO₂. Based on this observation, we know that the Martian surface became more oxidized with time. In fact, this increase in oxygen fugacity with time can also be inferred from samples of the Martian crust. The large fO₂ range in the shergottite meteorites from ~IW to ~5 log units above IW, coupled with their geochemistry, indicates that there is an oxidized crust-like component on Mars and a reduced mantle component [e.g., Wadhwa, 2001; Herd et al., 2002; Shearer et al., 2013; Steele et al., 2012].

The inferred range of fO₂ of the ~3.7 Ga Gusev basalts is similarly large but extends to more oxidized conditions [Schmidt et al., 2013; Tuff et al., 2013]. Finally, the most representative sample of Martian crust that we
have, NWA 7034/7533, assembled ~1.7 Ga, indicates that the Martian crust is highly oxidized [Humayun et al., 2013; Muttik et al., 2014; Santos et al., 2015]. The cause of this oxidation is likely related to the loss of H$_2$, which has an oxidative effect on the remaining system [Sharp et al., 2013], and the H$_2$/H$_2$O fugacity ratio of the atmosphere decreases, up to the point where the fO$_2$ defined by the H$_2$/H$_2$O fugacity ratio is equal to an oxygen fugacity that approximates the FMQ buffer (i.e., the point at which H$_2$ is exhausted from the system, equal to approximately IW + 3.5; see Sharp et al. [2013] for additional details). At FMQ, magmatic degassing favors the production of atmospheric species such as H$_2$O, CO$_2$, and N$_2$ over H$_2$. Although the change in redox evolution is understood and a probable mechanism identified, the timing of this change is much less certain and could have occurred as early as the waning stages of magma ocean crystallization or sometime after the cessation of the dynamo, where atmospheric loss was accelerated. Additionally, the relative controls of pressure and redox remain to be disentangled.

The rise of oxygen on our own planet is a complex feedback between the rise of cyanobacteria and their consequent changes on the geochemical cycles of redox-sensitive elements (e.g., C and P) [e.g., Kasting, 1993; Laasko and Schrag, 2014]. On Earth, recycling of carbon into the mantle may act to buffer the fO$_2$ of erupted lavas by reaction of carbonates and organic carbon. Mars is an opportunity to understand (presumably) abiotic oxidation, but where, as is typical for many possible exoplanetary cases [Harman et al., 2015], molecular oxygen in large quantities does not result. Losses in the Martian atmosphere through time of the lightest gas, molecular hydrogen, compound the magmatic effect and gradually lead to oxidation over time due to the buildup of elemental oxygen (i.e., electron-poor equivalents in the form of ferric iron-bearing minerals and sulfate salts) left behind from photolysis of water molecules [e.g., Hsurowitz et al., 2010]. It is not clear whether oxidative processes acting in Mars’ atmosphere and during weathering of its surface are detrimental to habitability; indeed, the redox potentials and energetic gradients between surface (oxidized) and subsurface (reduced iron phases in rocks and mantle) are increased. However, redox dictates the nature of gases present in the atmosphere, in turn governing nearly all aspects of water chemistry and climate.

3.3.3. How Do Volatile Species and Geochemical Cycles Control Climate?

The Tharsis volcanic load is composed of over $10^8$ km$^3$ of basaltic materials. Assuming magma water contents of a few tenths of percent to a few percent, this is equivalent to a global layer of water 10–100 m thick, which might have enabled the early warm and wet Martian climate, depending also on other volatile species present [e.g., Phillips et al., 2001]. Other components of atmospheric gases released from magma depend both on the composition of the mantle (starting materials; section 3.3.2), the composition of crustal materials (if these are metamorphosed to released gases), and the atmospheric pressure (section 3.1.2). Of course, which species are present influences the degree of greenhouse warming (section 2.6). Mars presents an opportunity to understand whether and why sufficient H$_2$ might be produced for early greenhouse warming but then cease [Ramirez et al., 2013], consequences of the speciation of sulfur on climate and geochemical cycles [Halevy et al., 2007; Johnson et al., 2009; Gaillard and Scaillet, 2014], the sequestration of carbon dioxide as carbonate [Niles et al., 2013; Edwards and Ehlmann, 2015; Wray et al., 2016], and the possible evolution of multiple water reservoirs [e.g., Usui et al., 2015]. As noted in section 3.1.1, Mars provides a key test case for inhibited volatile recycling. The rock record of mineral assemblages will record whether H$_2$, CH$_4$, or sulfurous species ever played a substantial role in greenhouse forcing and changes in redox and water availability through time.

3.4. A Late Climate Optimum? Effects of Faint Young Stars

Observations of exoplanets today are one snapshot of an evolving system. A primary parameter crucial to habitability of the planetary surface is the star itself, including the age of the star, which influences the history of processes causing atmospheric loss (sections 3.1.2, 3.2) as well as the overall history of incident radiant heating. For the latter, the “faint young Sun problem” has long been recognized as a challenge for explaining the existence of liquid water on Earth during its first billion years (for a recent review, see Feulner [2012]). According to standard solar models, the luminosity of the Sun upon ignition at ~4.57 Ga was ~30% fainter than at present (Figure 2). Because planetary equilibrium temperature is proportional to the 4th root of absorbed solar radiation, this implies equilibrium temperatures 5%–10% lower than today during the first billion years of planetary evolution, assuming only limited variation in the value of the planetary Bond albedo. The faint young Sun problem is especially acute for Mars, where ~70 K of greenhouse warming (more than twice that for Earth) [e.g., Wordsworth, 2016] is required to sustain liquid water, known from multiple sources of evidence to at least be episodically present at the surface (see section 2.5).
A late climate optimum was proposed as one explanation for the simultaneous occurrence of Al-phyllosilicates, weathered in open systems, and valley networks in the late Noachian, and their relative paucity in early Noachian units, while simultaneously noting a preservation bias might be equally probable [Ehlmann et al., 2011a]. For late Noachian Mars and even more so for earlier time periods, it is not yet clear from current climate modeling [e.g., Forget et al., 2013; Wordsworth et al., 2013, 2015] that atmospheric mass increases in pCO₂ alone can compensate for the presumed effects of a faint young Sun.

Solution(s) to the faint young Sun problem remains open. For Earth, at least four sets of solutions have been proposed: (1) increased concentrations of atmospheric greenhouse gases, supplied by hydrothermal activity or volcanism, including CO₂, H₂, CH₄, NH₃, N₂SO₂, and other species; (2) changes in the abundance, composition, and distribution of low- and high-level clouds, simultaneously changing planetary albedo and greenhouse warming; (3) changes in ocean heat transport due to less and differently distributed land masses and/or differences in ocean salinity; or (4) changes in axial obliquity and consequent impacts on heating and heat transport [Feulner, 2012].

Mars, with its long geologic record covering the early time period of increase in stellar flux provides a test case with an atmosphere of fundamentally different (arguably less complex) compositional evolution in which to untangle the changes in radiative forcing from different species. A northern ocean may have been present or absent, but significant changes in its configuration are unlikely, eliminating one variable (3). By sampling the geologic record, (1) and (4) can be directly assessed, looking for indicators of atmospheric species—largely not overprinted by diagenesis or life—and for periodicity of forcing.

3.5. Stochasticity and Impacts: Bringers of Destruction, Creation, or Both?

The assembly of any planetary system is chaotic and models of planet assembly and bombardment are probabilistic. Thus, the fate of any world with a given stellar flux, size, and composition is dependent on some aspects that are predictable and others that are stochastic. Key parameters are the timing of bombardment and magnitude. Yet these and their effects on planetary systems—did large impacts sterilize the planet? or generate a large gain or loss in volatiles?—can be modeled but are ill-constrained by data. Uniquely in the solar system, Mars preserves the ability to answer questions about the effects of impact bombardment on the habitability of a rocky body with an atmosphere.

3.5.1. What Were the Timing and Intensity of the Impact Bombardment Flux?

Multiple models for the time variation in the impactor flux to the inner solar system exist, resulting in two fundamental uncertainties: (1) the timing, duration, and magnitude of a putative late heavy bombardment and (2) timing of events younger than 3.5 Ga. For the first uncertainty, traditional crater counting models posit a smoothly declining bombardment rate with a half-life defined by the end of the accretionary period [e.g., Hartmann and Neukum, 2001]. Based on lunar impact melt dating, many other models contain a “spike” in bombardment in the inner solar system around 3.9–4.1 Ga, though the detailed impact flux and peak time vary. An increase in bombardment flux might result from the dynamical interplay between the giant planets, the primordial disk, and the myriad of smaller bodies in the asteroid and Kuiper belts. Some models favor a narrow spike late heavy bombardment at 3.9 Ga, justified by age dating of lunar materials [Ryder, 2002], while other models favor a more distributed sawtooth-like heavy bombardment period with a peak near 4.1 Ga [Morbidelli et al., 2012a].

For the second uncertainty, recent work argues for a range of lunar chronologies [e.g., Hartmann et al., 2007; Marchi et al., 2009], with differences between the models of up to 1 billion years [Robbins, 2014]. These differences are primarily due to different assessments of crater densities, especially in recent Lunar Reconnaissance Orbiter data, and a lack of dated samples of known provenance for terrains with N(1) crater densities of ~0.0015 km⁻² to 0.0025 km⁻². This seemingly small range is associated with the time period from ~1–3.5 Ga, with the largest differences between the models occurring at about 3.3 Ga. Mars, the Moon, and possibly Mercury provide the main records of this impactor flux as Venus and the Earth have had their records overprinted by later volcanism, tectonism, and erosion. The record in the asteroid belt is complicated by intra-belt dynamics. The lunar record is certainly the most pristine, since early surfaces on both Mercury [Marchi et al., 2014] and Mars [Tanaka, 1986] experienced resurfacing by volcanism and erosion concomitant with heavy bombardment. Indeed, the number of the largest Martian basins is debated (see section 2.2), a fact which complicates chronologies. Nonetheless, Mars provides a key test of our understanding of solar system chronology. Do the extrapolations of the impactor flux and timing based on the age and distribution of lunar...
record, terrestrial isotopes, and known physics also explain the Mars chronology? Determining ages for ancient Martian impact basins and early volcanic provinces would definitively constrain the timing of inner solar system heavy bombardment and its effects on the evolution of potentially habitable worlds.

3.5.2. Do Early Large Impacts Sterilize a Planetary Surface?
The impact flux of the Earth-Moon system remains only partially constrained and a nearly 2 orders of magnitude difference in the impactor flux as well as differences in the impactor size-frequency distribution dictate whether the Hadean Earth experienced overall moderate but very locally severe impact heating [Abramov and Moizsis, 2009] or several planet-wide near-sterilization episodes [Marchi et al., 2014]. The Martian record from this period exists, though it is not presently clear whether layered strata represent impact breccias, volcanic lavas and pyroclastics [McEwen et al., 1999; Bandfield et al., 2013], or sedimentary deposits. Dominance of melts and breccia might imply episodic sterilization; dominance of sedimentary and volcanic environments might instead imply “normal” geologic processes persisted. In situ petrology can distinguish the origin of Mars’ earliest units and isotopic and chemical data from mineral phases can establish surface temperatures.

3.5.3. A Net Subtraction or Addition of Atmospheric Volatiles?
A key area of debate is the role of impacts on atmospheric loss. Net addition or removal depends on the mass of the planet, the mass of the atmosphere, and the size, velocity, and composition of the impactors. Atmospheric loss by large impacts was modeled to be significant [Melosh and Vickery, 1989] with projectiles as small as 3 km in diameter proposed to cause Martian atmospheric loss and complete atmospheric removal for a 160 km diameter projectile at 20 km/s [Ahrens, 1993]. Recent work has focused attention on observed enrichment of radiogenic $^{40}$Ar relative to $^{36}$Ar, which would have been removed from impacts, and the efficiency of small impactors in atmospheric removal with most recent models supporting considerable impact erosion [Manning et al., 2007; Svetsov, 2007; Pham et al., 2011; Schlichting et al., 2015]. Calculations of atmospheric escape per impact have in other cases been revised downward, including net mass addition for projectiles up to 30 km diameter [Newman et al., 1999; Shuvalov and Artemieva, 2002] and for the cumulative bombardment flux at Mars [de Niem et al., 2012]. Isotopic measurements of trapped Martian paleoatmosphere are the means to resolve this question by constraining the relative importance of different volatile sources. Direct atmospheric pressure paleoproxies are challenging in the geologic record; nevertheless, small compositional changes due to loss of atmosphere or addition of impactor volatiles may be discernible in Martian materials before and after formation of its large basins.

3.6. Cyclical Habitability: The Role of Obliquity, Eccentricity, and Rotation Rate?
On Earth, with its continuous supply of liquid water, nutrients are a typical limit on life, rather than water itself. On Mars, it seems clear that habitability varied as a function of location and time much more acutely due to actual limitations on water. Indeed, at the Mars surface at present liquid water is only stable against sublimation for ≤10% of a Martian year over <30% of the surface [Haberle et al., 2001]. However, changes in orbital parameters mean this limitation was not always so acute.

Mars undergoes changes in the latitudinal distribution of solar insolation, driven by chaotic obliquity, i.e., changes in axial tilt, over timescales of ~125 kyr [Ward, 1974]. Though Mars’ present-day obliquity of 25° is similar to that of Earth’s (23.3°), such has not commonly been the case. In just the last 20 Myr, Mars’ obliquity has fluctuated from 15 to 45°, and it has likely exceeded 60° within 100 Myr to 1 Gyr [Laskar et al., 2004], making it the most variable in the modern, inner solar system (Figure 7). These profound changes in Martian obliquity, along with changes in eccentricity, greatly influence the stability of volatiles on the Martian surface. At low obliquities (<20°), the Mars atmosphere can collapse to surface pressures of just a few millibars [Leighton and Murray, 1966; Soto et al., 2015]. During higher obliquities (>45°) glaciers and water ice sheets can form at midlatitudes [Forget et al., 2006], and under only moderately higher pressures, temperatures are sufficient to melt ice during part of the year [Wordsworth et al., 2013]. Changes in planetary orbital eccentricity also play a role in climate forcing.

In contrast, paleomagnetic studies indicate stability in Earth’s obliquity (±1.3°) at least since Archean time [Evans, 2006] and most likely earlier [Laskar and Robutel, 1993]. This orbital parameter was locked in with a (stochastic) giant impact and formation of a large satellite, which Mars lacks. Venus and Mercury once had chaotic obliquities similar to that of Mars today but are now stabilized by the Sun at very small, almost constant obliquities [Goldreich and Peale, 1970; Laskar and Robutel, 1993; Van Hoolst, 2015]
Thus, a key question is whether Mars’ extreme obliquity-driven “duty cycling,” superimposed on any secular trends in surface conditions, has been harmful or helpful to habitability. Similarly, one might consider the day-night “duty cycle,” of similar length for Earth and Mars (at least today) but potentially of long duration on tidally locked planets and certainly of long duration on Venus. Do cycles of short duration helpfully refresh while longer cycles harmfully disrupt? Or is the effect the reverse or neutral? On Mars, obliquity may have been crucial in episodically generating liquid water by providing a means to place water ice near the equator, where it would be more susceptible than polar ice to melting in the presence of other temporary climate perturbations (e.g., from volcanism, impact). It may also be the case that these perturbations made the surface an unstable habitat, while the subsurface habitats persisted relatively unaffected. The full environmental effects of Mars’ profound climate cycles and their impact on making habitability intermittent or continuous must still be explored.

4. Mars as a Linchpin for Terrestrial Evolution: Requirements for Exploration

Mars formed early and thus possesses a truly ancient record of the early planetary processes described above. Over 60% of the surface of Mars dates from the first billion years, and select stratigraphic sections (Figure 3; e.g., Mawrth Vallis, Nili Fossae, NE Syrtis) extend even earlier, prior to 3.9 Ga, presenting an opportunity to access an intact record from the pre-Noachian. The rocks of ancient Mars record the earliest planetary crusts, loss of the magnetic field, early intense impact bombardment, widespread volcanism, and the effects of these on the atmosphere, hydrosphere, and evolution of habitability. Yet though this environmental record is preserved, it has not yet been interrogated at a level that the nature of environmental change can be understood. From experience with the terrestrial record, many of the outstanding questions above can in fact be addressed with a richer suite of data sets. We outline those measurements required—all of which use technologies available—below. Note that many valuable investigations focused on modern dynamics or the surface and atmosphere will not appear; this is not a commentary on their importance but rather a reflection of the focus of this review (see also section 4.5). To interpret the evolution of Mars as a linchpin for understanding solar system history and controls on terrestrial planet evolution, the required measurements are summarized into four thematic categories (Table 3)—petrology, history of volatiles, absolute dating, and geophysical variability—each of which has specific measurements identified (Table 4).

4.1. Petrology at Multiple Sites: What Environments When and Where?

Understanding the origin and history of the Martian rock units under consideration is essential for contextualizing and interpreting geophysical, chemical, mineralogical, or isotopic data from them. It also holds a key record of planetary change between accretion and the apparent decline in liquid water availability ~3.5 Ga.
## Table 3. Necessary Measurements for Understanding the Evolution of Mars Habitability and Their Means of Acquisition: Orbital/Airborne, In Situ, or Sample Return Measurements

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Orbit/Air</th>
<th>In Situ</th>
<th>Sample Return</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stratigraphy and petrology measurements</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nature of the early Mars rock record, i.e., relative dominance of igneous, impactite, sedimentary units, and record in subcentimeter textures</td>
<td>partial</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Clay formation environment(s) (weathering, diagenetic, hydrothermal, and metasomatic/deuteric)</td>
<td>partial</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Identify marker beds in key stratigraphies</td>
<td>partial</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Volatile geochemistry, cycling and loss</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H, C, S, N, and O mineral phases in rocks and soils as a function of time and setting</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Gas/liquid inclusions in quench melts as a function of time</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Redox indicator minerals and phases, e.g., siderite, Mn phases, and S8, as a function of time</td>
<td>partial</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Multiple S and O isotopes for redox as a function of time</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Clumped isotopes for water temperatures as a function of time and setting</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Age date polar caps to understand role of orbital forcing in climate (possibly via ash layers)</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paleopressure from vesicles</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time-varying high-resolution gravity (seasonal mass flux for ice and water)</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Timing and effects of key processes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timing and duration of heavy bombardment from basin ages and melt ages</td>
<td>partial</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Time-pinned correlative stratigraphy distinguish local from global trends (age dating and marker beds)</td>
<td>partial</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Environmental effects of large volcanic eruptions and large impacts <em>via</em> age dating of volcanic/impactite rocks (and cross correlation with stratigraphies)</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Recognition of unconformities to understand environmental (dis)continuity</td>
<td>partial</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Timing and duration of valley networks from cross-cutting relationships <em>and</em> age dating</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Geophysical evolution</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetic field anomaly locations</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magnetic field strength <em>as a function of time</em></td>
<td>partial</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Igneous mineral geochemistry for mantle temperature <em>as a function of time</em></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Igneous geochemistry mantle redox state via minor and trace elements, e.g., V/Sc, Cr <em>as a function of time</em></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>REE to understand magma ocean and differentiation evolution</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interior structure from seismology</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spatially distributed heat flux</td>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aWhen measurements must be time-pinned samples, italics indicate the need for age dates. “Pos.?” means the measurement is possible from the given type of platform and “Req.?” means it must be made from that platform.*
Table 4. Key Needed Measurements

<table>
<thead>
<tr>
<th>Measurement Needed</th>
<th>Measurement Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orbital/Airborne</td>
<td></td>
</tr>
<tr>
<td>Magnetic field strength</td>
<td>Global spatial coverage at tens of kilometers sampling</td>
</tr>
<tr>
<td>Water reservoirs: extent and volume of water ice in nonpolar regions; subsurface capacity to store water</td>
<td>Radar and thermal inertia to quantify near-surface ice deposits and buried fluvial morphologies at spatial scale of 20 m and vertical sensing depths to 20 m; high-resolution gravity to determine porosity and capacity for pore waters, aquifers, ice, and temporal changes in the subsurface water reservoir</td>
</tr>
<tr>
<td>Mineral assemblages</td>
<td>SWIR and MIR spectroscopy to recognize primary and secondary mineral phases at 5% abundance and semiquantitative modal mineralogy of assemblages at ~10 m/pixel</td>
</tr>
<tr>
<td>Compositional stratigraphies</td>
<td>Imaging at tens of centimeters for morphologic context over ~5–10% of Mars; spectral imaging or multispectral color in SWIR and MIR at &lt;5 m/pixel resolution of ~20% of Mars</td>
</tr>
<tr>
<td>In situ (or Sample Return)a</td>
<td></td>
</tr>
<tr>
<td>Petrology (mineralogy + texture)</td>
<td>Image microscopic (tens of micrometers) textures of rocks, including coupled compositional data to identify minerals (±1%) and their phase relationships</td>
</tr>
<tr>
<td>Sample in situ age from radiometric isotopes</td>
<td>Precision of ±100 Myr on key lithologies representing Martian surface epochs; geochemical and geological context. Most plausibly done with measurement of lavas or volcanic ashes, with parent and daughter isotopes each measured with few percent precision</td>
</tr>
<tr>
<td>Electronic structure, coordination, and redox</td>
<td>Identify and quantify Fe(III)/Fe(T) for bulk sample, separating by phase if possible; Ni, Cr abundances in bulk or phases; X-ray absorption spectroscopy for Fe, Mn, S</td>
</tr>
<tr>
<td>H, C, S, Cl, and N-bearing mineral phases</td>
<td>Identify and quantify quantity at ±1%; this includes ID of mineral species and major cations</td>
</tr>
<tr>
<td>H, C, S, N, and Cl bulk isotopes</td>
<td>Isotope ratios D/H, 13C/12C, 15N/14N, 34S/32S, and 37Cl/35Cl to &lt;0.1‰ in H2O, CO2, SO2, and HCl in solid samples</td>
</tr>
<tr>
<td>Heat flux</td>
<td>Determine surface heat flux ±2 W/m² at multiple locations (~3) to understand present-day heterogeneity (if any)</td>
</tr>
<tr>
<td>Interior Structure</td>
<td>Seismometry at sufficient station numbers to determine core-mantle boundary depth, crustal thickness, and presence/absence of failed slabs</td>
</tr>
<tr>
<td>Sample Return</td>
<td></td>
</tr>
<tr>
<td>H, C, N, O, Cl, and S phases and isotopes in mineral phases and in fluid inclusions</td>
<td>D/H, 13C/12C, 15N/14N, 18O/16O, 34S/32S, and 37Cl/35Cl in different mineral phases</td>
</tr>
<tr>
<td>Multiple isotopes</td>
<td>34S/32S, 32S and 18O/16O to &lt;0.1‰ in a variety of volatiles in returned samples</td>
</tr>
<tr>
<td>Noble gas isotopes in rocks</td>
<td>He, Ne, Ar, Kr, and Xe isotopic composition to &lt;2%</td>
</tr>
<tr>
<td>Clumped isotopes for water temperatures</td>
<td>Clumped isotope measurements of carbonates and organics</td>
</tr>
<tr>
<td>Metal isotopes</td>
<td>Key systems for redox history include Mo, Cr, and Fe</td>
</tr>
<tr>
<td>Trace elements in bulk rock and discrete phases</td>
<td>Measure elements at the sub-10 ppm level in rock and constituent rock-forming minerals</td>
</tr>
<tr>
<td>Sample age from radiometric isotopes, by mineral phase</td>
<td>Measure isotopes of main radiogenic systems (e.g., U-Pb, Rb-Sr, Sm-Nd, and Lu-Hf) at the sub parts per million level in mineral separates and/or in suitable phases (e.g., U-Pb in zircon or baddeleyite)</td>
</tr>
<tr>
<td>Magnetic field strength and orientation</td>
<td>Acquire information on the intensity and the direction of ancient fields</td>
</tr>
</tbody>
</table>

Note some of the in situ measurements can be done on returned samples—and done better—if their context is clear and if sufficient numbers of samples from multiple sites are returned so as to understand their representativeness for global change (see text).

The nature of many younger rock units exposed at the surface is self-evident as they are frozen in time, preserving much of the morphology of their emplacement. However, the most ancient Martian record is considerably more difficult to decipher and has confounded orbital analyses to date. Units are coherent but typically exposed only in vertical section in select locations where faulting or erosion have revealed mid-Noachian and older materials buried beneath younger surfaces (e.g., Figures 2 and 3). The present best orbital resolution for interrogating the nature of these units is 0.3 m/pixel (provided by High Resolution Imaging Science Experiment) and available for ~2% of the Martian surface as of this writing, and no lander/rover ever explored such ancient crust sections. In many ancient, buried units, there are clay or carbonate minerals and the stratigraphy is layered, but the length scale over which the layers are continuous is only hundreds of meters [Mustard et al., 2009; Michalski and Noe Dobrea, 2007; Bandfield et al., 2013; Stack et al., 2013]. Are these volcanic flows, pyroclastics, sedimentary units (fluvial, lacustrine, glacial), or impactites? Or some combination? These units are also typically clay mineral-bearing and thus represent a reservoir of rock-sequestered H2O, CO2, SO2, and HCl – 0.1‰ in a variety of volatiles in returned samples.
postdepositional alteration by rain or snow melt [e.g., Carter et al., 2015] (see section 2.4); however, the timing of the alteration relative to emplacement is currently unknown so constraining these models is challenging given current orbital instrumentation. Do the clay minerals record early weathering, diagenetic or hydrothermal alteration in Martain aquifers, or direct crystallization products from magma [Meunier et al., 2012]? The question is important for understanding planetary habitats because these units preserve the record of environmental conditions during what may have been the most habitable period on Mars while the magnetic field was active, the atmosphere thicker, and volcanic and impact processes vigorous. But the nature of the habitats is unknown.

The essential measurements required to understand the nature of the early hydrated units are centimeter-to subcentimeter-scale petrologic analyses, i.e., coanalysis of rock textures coupled with mineralogy of multiple units in stratigraphy (Table 3). These observations will determine whether predominant environments on earliest Mars were volcanic, sedimentary, or impact-driven and the nature of the surface- and near-surface environment. Specifically relevant to habitability, the measurements of rock chemistry and mineralogy can determine the salinity, oxidation state, and availability of key nutrients in surface and subsurface habitats with liquid water. Petrologic observations will also determine how later geologic processes such as erosion, impact heating, aqueous alteration, and oxidation overprint (or not) the original lithology and contribute to the signatures from orbital data, influencing interpretations. The subcentimeter spatial scale required necessarily demands that measurements be made or samples collected in situ, with the ability of nanometer- to few micrometer-scale data in returned samples to provide still more refined petrologic information. Of course, equivalent data will not be available for exoplanets. But it is this data from Earth and from Mars that permit assessing changes in habitability with time, what processes drove their change, and finally, how signatures of the driving processes might be recognized remotely.

4.2. Follow the Volatiles to Track Geochemical and Climate Change: Inventories, Isotopes, and Submeter Scale Stratigraphies

A planet’s atmospheric pressure and compositional evolution depend on every parameter in Table 1. The only way to disentangle the tightly coupled processes in such a system is to develop geochemical models, constrained by reservoir sizes, geologic fluxes, surface physical and radiation environments, and loss to space. Though uncertainties remain, much progress has been made in characterizing the Martian volatile reservoirs for water ice [Lasue et al., 2013, and references therein], CO2 ice [Phillips et al., 2011], carbonates [Niles et al., 2013; Edwards and Ehlmann, 2015; Wray et al., 2016], nitrates [Stern et al., 2015], oxychlorine compounds [e.g., Archer et al., 2014; Ming et al., 2013], sulfates [Gaillard et al., 2013, and references therein; McAdam et al., 2014], and trace organic compounds detected in the atmosphere [Webster et al., 2015] or thermally released from ancient rocks [Freissinet et al., 2015]. The MAVEN mission is presently characterizing the processes and rates of modern atmospheric loss [Jakosky et al., 2015b, 2015c; Brain et al., 2015], and the Mars Science Laboratory (MSL) mission is characterizing both atmospheric and surface volatiles and searching for organic compounds to understand the sources and preservation of reduced molecular compounds in this environment.

The largest remaining questions are evolutionary: what was the timing of changes in atmospheric composition, redox, and pressure? On what timescales do surface and subsurface reservoirs exchange? Which atmospheric loss processes have dominated and why? How do radiation processes impact the lifetime of complex organic molecules in the near-surface environment [Pavlov et al., 2012]? The required observations are at spatial scales that subsample rocks or sample the mineral grains within rocks. New data from MSL of light isotopes in the atmosphere [Webster et al., 2013; Mahaffy et al., 2013] and in bulk materials [Leshin et al., 2013; Mahaffy et al., 2015] have opened up a new means of inquiry for understanding the fate of volatiles, complementing the record from meteorites [Bogard et al., 2001; Filiberto et al., 2016]. Bulk or spatially resolved chemistry and mineralogy refine the C, H, N S inventories as well as suggest loss processes with time. For example, D/H changes over time are complicated by the uncertainty in the location and isolation of near-surface H2O reservoirs with different isotopic compositions, but in the simplest scenarios the D/H ratio has changed monotonically over time due to water loss to the atmosphere [e.g., Mahaffy et al., 2015; Villanueva et al., 2015]. An alternative is that multiple reservoirs have been created, including an intermediate crustal reservoir of ice or minerals [Usui et al., 2015]. Key measurements for hydrogen and all other volatiles relevant to atmospheric composition are to track the type and isotopic signature of secondary minerals formed at different times. Isotopic measurements of H, C, S, and N isotopes,
especially when in petrologic context so that fractionation during sequestration can be understood, provide a record of atmospheric isotopic fraction due to various loss processes. Distinct phases and multiple isotope systems indicate redox chemistry. Direct atmospheric samples are also likely available in many quenched rocks from impacts or volcanism, and trapped noble gases play a key role in tracking loss and addition. Thus, fundamental questions about controls on volatile reservoirs, cycling, and atmospheric composition can be directly addressed with in situ or sample return mineralogical and isotopic measurements.

These measurements also feed forward to understanding fundamental controls on atmospheric loss, disentangling size, distance from Sun, magnetic shielding, and starting composition—measurable or predictable by models for distant planetary systems—from more stochastic and/or derived parameters such as impact history and the timing of volatile release. By interrogating the isotopic record of volatile-bearing rocks before and after the loss of Mars’ magnetic field, the effects of sputtering may be distinguished from other processes. By interrogating the record before and after giant (>500 km) impacts, the role of impacts as net removers or deliverers of volatiles can be assessed.

4.3. Timing Is Everything: Understanding of Processes on Mars and Across the Inner Solar System by Measuring Ages

Mars can serve as a critical archive of inner solar system chronology because it preserves a record of the impactor flux history and also, possibly, changes in stellar activity that drove atmospheric change. The vast majority of Martian samples (meteorites) unfortunately contain little evidence for impact resetting during their residence on Mars, but we also know from their peculiarly young age distribution that these meteorites do not represent most exposed terrains on the surface of Mars. Radiometric dating of returned samples is the most straightforward way to identify formation ages as well as overprinting episodes associated with major impacts, but recent advances have now also demonstrated in situ age dating [Farley et al., 2013b]. Multiple instruments designed and optimized specifically for age dating have been developed for flight [Farley et al., 2013a; Cohen et al., 2014; Solé, 2014; Anderson et al., 2015a, 2015b; Cho et al., 2016] opening up the possibility for missions to carry payload elements designed to both elucidate the Martian planetary chronology and give solar system-wide context to Mars’ geochemical change with stratigraphy as it is explored.

Determining the existence and temporal evolution of any late heavy bombardment episode is a particularly critical measurement. Reliable radiometric ages from ancient sites where the cratering record is clear would also help anchor the cratering chronology and help constrain the timing of transitions in Mars’ magnetic field, mineralogy, and environment. Mars is the only terrestrial planet where this is possible, so this type of measurement would establish context for early evolution of the Earth and Venus as well. Critical questions include: Did Mars experience an increased flux in large impacts at ~4.1–3.9 Ga, as many people think occurred on the Moon, Earth, and asteroid belt, or did Mars instead experience a simple exponential decline in bombardment? Was the timing of flux at the Moon and Mars the same?

Orders of magnitude of uncertainty presently exist in the rates of Martian depositional and erosional processes, perhaps most acutely for the valley networks and sedimentary deposits. Use of cross-cutting units (impactites, volcanic ashes or lavas) to date distinct stages of fluvial activity and sedimentary deposition could resolve longstanding questions of episodicity or continuity of surface waters. Accumulation rates and erosion rates can be constrained using craters [e.g., Kite et al., 2013b] and cosmogenic nuclides [e.g., Farley et al., 2013b]. Disentangling global versus local environmental change requires time-pinned correlative stratigraphy to understand whether similar shifts are observed elsewhere on the planet as a function of time. Coupled with orbital data, this may lead to the identification of specific features, e.g., marker beds, which can be observed globally and used to correlate sections. Equally important is the identification of unconformities and quantifications of gaps in the record to understand the continuity or abruptness of environmental change.

Crucially, correlation of isotopic and chemical records of the atmosphere with, e.g., volcanic eruptions or large impacts allows answering questions little understood across solar system bodies. What volatiles were emitted during volcanism and did they warm or cool the climate? Do impacts add or remove volatiles? With stratigraphy coupled to absolute dating, these processes can be determined using key Martian sections with igneous and sedimentary units. Absolute dating to cross-correlate and establish timing of planetary response to major events is critical.
4.4. The Coupling of Interior and Surface Evolution

The Martian interior controls the hydrosphere and atmosphere profoundly but in a way such that the couplings between these systems as yet remain difficult to define, especially extending further into the past. The InSight mission [Banerdt et al., 2013] is expected to measure heat flux in one location and by single station seismometry, characterizing in part the interior structure. Indeed, these are crucial measurements which can be extended by measurements of heat flow in other locations to assess whether this is heterogeneous or homogenous for a planet with Mars’ style of mantle circulation. Additional seismometry data may also be needed to understand the nature of the core, location of the core-mantle boundary, crustal thickness, and the existence of any failed slabs that might signify prototectonics. Igneous minerals and bulk rock chemistry can trace the mantle potential temperature through time, and rare earth elements (REEs) and minor elements provide the clues for magma ocean existence as well as mantle redox state and change.

Other key questions include the exact timing and duration of the magnetic field shutoff, and whether a second, restarted dynamo ever existed. When the field initiated and declined, whether underwent multiple starts and stops, and the effect of the magnetic field on atmospheric loss are all recorded in pre-Noachian to early Noachian rocks.

5. Future Outlook

5.1. Future Mars Exploration Approaches

Mars has now been visited by a plethora of orbiters: what remains for exploration from above? “Above” includes orbiters as well as newly developed airborne platforms including Mars helicopters or balloons. There remain many key measurements for Mars orbiter/airborne instruments for understanding the modern, dynamic Mars environment and for landing site geologic and safety characterization (see Mars Exploration Program Analysis Group-Next Orbiter Science Analysis Group Report [2015] and MEPAG [2015] supporting information tables for a summary of key measurements for modern Mars). For example, the Mars Exploration Program Goals Document [MEPAG, 2015] states, “…an understanding of the present climate must be firmly established before an understanding of the climate of the recent past can be developed. Numerical models play a critical role in interpreting the recent past and ancient climate, and it is imperative that they be validated against the present climate in order to provide confidence in results for more ancient climates that are no longer directly observable.” Thus, for understanding long-term evolution, the key orbiter or airborne measurements remaining relate to understanding the current climate, and mapping at higher spatial resolution the magnetic anomalies along with additional data to resolve the bedding and composition of stratigraphic sequences (section 4.2; Table 3).

What is remarkable in Table 3 is how much progress on key questions in Mars early history can be accomplished from landed payloads. The traditional tasks of landed missions are mesoscale and microscale imaging, bulk mineralogy, and bulk chemistry. However, many categories of investigation previously in the sample return category have migrated to in situ instrumentation as instrument capabilities improve, and this trend will continue. Within the last decade a host of instrument and sample handling techniques for doing in situ submillimeter petrology and in situ isotopic analyses of rocks and sediments for stable isotopes and age dating have become available. Coupled to mobility, these effectively enable stratigraphic columns to be fully characterized in situ. Of course, sample return remains important as submicrometer imaging with composition to resolve phase differences (comparable to SEM-EDS, microprobe, and synchrotron studies) and many clumped isotope or other isotopic and REE systems remain out of reach on a multidecadal time period for anything but terrestrial laboratories whose analytical capabilities too continue to increase.

Because of the complexity of the Mars system and the different time periods and environments available at different geographical locations, single site sample return is insufficient to address the questions posed here. Rather, as on Earth, interrogations of multiple stratigraphies are required to capture the most pressing science questions. The need to disentangle local unique environments from truly global change is fundamentally a question of sampling. The cost and programmatic effectiveness of several in situ missions with highly capable rovers or several sample return missions, possibly with less capable rovers, is a key programmatic choice. This must balance capabilities and resources available. However, importantly, Table 3 demonstrates there is significant progress that can be made in understanding early Martian history and, indeed early solar
system history and the key controls on terrestrial planet evolution and habitability, by in situ as well as sample return missions.

5.2. Future Rocky Exoplanet Exploration Approaches

Many of the parameters governing terrestrial evolution and habitability (Table 1) can be determined telescopically for exoplanets. Using new data from exoplanets to answer questions about the parameters controlling the sustainability of habitability (Table 2) is more challenging. Much of the data to ground truth models applied to planetary systems around other stars will come from the record in our own solar system. Nonetheless, data from exoplanets (e.g., temperature, presence/absence of an atmosphere, and its composition) could certainly aid our understanding of the effects of orbital parameters on habitability (#6, Table 2), the effects of stellar evolution/age on planetary climate (#4), and information on how planetary size and long-term atmospheric loss/composition are coupled.

The near-term future entails first detecting more and smaller rocky exoplanets. For transiting exoplanets around a Sun-like star, Earth- and Mars-sized planets result in brightness decreases of 84 ppm and 24 ppm, respectively. The depth of the transit is used to determine the radius of the orbiting planet. The masses of both transiting and non-transiting planets can be determined via the radial velocity technique (the “Doppler wobble”) and used with the inclination $i$ of the system to convert the measured signal into the true planet mass. In our solar system, Jupiter induces a Doppler wobble of 12.5 m/s, the Earth causes a 9 cm/s wobble, and Mars generates a tiny motion of 5 cm/s. Mars-like planets will be challenging to detect around active, Sun-sized and larger stars. But a Mars-sized planet orbiting within the habitable zone of a small red dwarf would generate a Doppler wobble of roughly 40 cm/s. Current exoplanet surveys are sensitive to transit depths of tens of parts per million and radial velocity semi-amplitudes of roughly 50 cm/s. The next generation of radial velocity spectrographs aspire to measure signals smaller than 10 cm/s. For example, the Giant Magellan Telescope-Consortium Large Earth Finder has a precision goal of $\sim$10 cm/s [Szenti-Gyorgyi et al., 2016]. Additionally, the Eschelle SPectrograph for Rocky Exoplanet and Stable Spectroscopic Observations on the European Southern Observatory’s Very Large Telescope array should have a precision of radial velocity variations of a few centimeters per second [Pepe et al., 2014]. Thus, a continuing expansion of the number of rocky exoplanets and their basic order characteristics—size, density, and orbital position—seems likely to continue to the foreseeable future.

As this paper was being written, astronomers announced the discovery of a probable rocky planet around the nearest star, Proxima Centauri [Anglada-Escudé et al., 2016]. This planet orbits at $\sim$0.05 AU and has a minimum mass of 1.3 Earth masses. The star, of spectral type M5.5 V, has a mass of only 0.12 that of the Sun, so the radial velocity signal is relatively large, $\sim$1.4 m/s. The luminosity of the star is $\sim$0.15 that of the Sun; hence, the stellar flux received by the planet is $\sim$0.6 that at Earth and 1.3 that at Mars, putting it well within a conventional liquid-water habitable zone [Kopparapu et al., 2013]. The planet does not transit, so it is difficult to observe using present space telescopes, but it could eventually be observed from the ground with planned 30–40 m telescopes equipped with adaptive optics and high-contrast coronagraphs.

The knowledge that will vastly broaden understanding of the habitability and evolution of these rocky worlds is information on composition of an atmosphere or surface, probably obtainable within a timescale of decades. Broadband photometry of secondary eclipses in MIR emission would be able to determine whether a given planetary surface is of silicates or iron oxides at precisions of 2 ppm for planets around G stars and 20 ppm for planets around M stars, if any atmosphere is relatively transparent [Hu et al., 2012]. This is at the threshold of presently planned measurement capabilities for transit observations of spectrophotometric properties of Earth-sized planets [e.g., Beichman et al., 2014]. Few to tens of parts per million precisions are sufficient for detection of certain gases species, e.g., CO$_2$, but quantification would require a 2–5 times improvement in signal-to-noise ratio [von Paris et al., 2013]. Instruments for doing so are under study, including the High Resolution Spectrograph for the European Extremely Large Telescope [Marconi et al., 2016]. Reflected stellar light can also be used to uniquely determine the mineral composition of an exoplanet’s surface. However, at $<0.1$ ppm for an Earth-sized exoplanet around a Sun-like star and 1 ppm for an Earth-sized exoplanet around an M dwarf, it is out of reach for transit observations due to stellar variability [Hu et al., 2012]. Eventually, the unique identification of minerals on exoplanets’ surfaces may rely upon direct imaging, and degeneracies between surface- and atmospheric-related features may be broken by achieving high spectral resolution [Hu et al., 2012]. Armed with the majority of the parameters in Table 1 and a vastly...
improved understanding of the questions in Table 2, including predictive models, the habitability of these
to-be-discovered terrestrial planets can be assessed.

6. Conclusions

Ideally, the best way to determine controls on planetary habitability would be a series of isolated experiments:
what would happen if Earth were moved to 1.5 AU and Mars to 1 AU? What if Earth had no Moon and experi-
enced chaotic obliquity excursions? What if Mars had a magnetic field throughout its history and Earth did
not? Of course, these types of questions are limited to modeling exercises only. The only way to disentangle
these parameters is historical study, e.g., geologic fieldwork and compositional measurements (at scales from
regional to microscopic), coupled with modeling with assumptions constrained by observation. This can only
be accomplished if data recording relevant processes are available, and these data can be obscured, modi-
ﬁed, or erased by subsequent processes. Such is the challenge and the opportunity for examining the geolo-
gic record of early Mars. Mars has a rich, complex, but organized geologic record of fundamental processes,
certainly extending back past 4 Ga and possibly to as long ago as a mere 5 Myr after solar system formation in
the isotopic fingerprints of its initiation as a planetary embryo. A rich suite of questions about the long-term
sustainability of habitability is addressable through future Mars measurements, most notably in situ petro-
logy, stratigraphy, and measurement of isotopes and absolute ages coupled with sample return to interrogate
the time-ordered record of early processes at even ﬁner scale. The diversity of Mars geologic history demands
interrogation of multiple stratigraphies and time periods to build the required knowledge. Planet formation
models need to robustly incorporate geochemistry/volatiles/isotopic signatures to consider the hypothetical
habitability of exoplanets formed under very different system architectures than our own. Armed with our
understanding of the divergent habitability paths of Earth, Venus, and Mars, we can look outward to the array
of terrestrial exoplanets to be discovered to understand their potential to host and sustain life.

Acknowledgments

One of us (Y.L.Y.) thanks Vlada
Stamenkovic for illuminating
discussions of the Martian atmosphere
at the KISS Workshop on Methane on
Mars. As per AGU’s data availability
policy, this paper is a review paper and
contains no new data. Thanks to two
reviewers and the editor for comments
that improved this manuscript.

References

Acuña, M. H., et al. (1999), Global distribution of crustal magnetization discovered by the Mars Global Surveyor MAG/ER experiment, Science, 284, 790–793.
sustainability of habitability is addressable through future Mars measurements, most notably in situ petro-
logy, stratigraphy, and measurement of isotopes and absolute ages coupled with sample return to interrogate
the time-ordered record of early processes at even finer scale. The diversity of Mars geologic history demands
interrogation of multiple stratigraphies and time periods to build the required knowledge. Planet formation
models need to robustly incorporate geochemistry/volatiles/isotopic signatures to consider the hypothetical
habitability of exoplanets formed under very different system architectures than our own. Armed with our
understanding of the divergent habitability paths of Earth, Venus, and Mars, we can look outward to the array
of terrestrial exoplanets to be discovered to understand their potential to host and sustain life.

6. Conclusions

Ideally, the best way to determine controls on planetary habitability would be a series of isolated experiments:
what would happen if Earth were moved to 1.5 AU and Mars to 1 AU? What if Earth had no Moon and experi-
enced chaotic obliquity excursions? What if Mars had a magnetic field throughout its history and Earth did
not? Of course, these types of questions are limited to modeling exercises only. The only way to disentangle
these parameters is historical study, e.g., geologic fieldwork and compositional measurements (at scales from
regional to microscopic), coupled with modeling with assumptions constrained by observation. This can only
be accomplished if data recording relevant processes are available, and these data can be obscured, modi-
ﬁed, or erased by subsequent processes. Such is the challenge and the opportunity for examining the geolo-
gic record of early Mars. Mars has a rich, complex, but organized geologic record of fundamental processes,
certainly extending back past 4 Ga and possibly to as long ago as a mere 5 Myr after solar system formation in
the isotopic fingerprints of its initiation as a planetary embryo. A rich suite of questions about the long-term
sustainability of habitability is addressable through future Mars measurements, most notably in situ petro-
logy, stratigraphy, and measurement of isotopes and absolute ages coupled with sample return to interrogate
the time-ordered record of early processes at even finer scale. The diversity of Mars geologic history demands
interrogation of multiple stratigraphies and time periods to build the required knowledge. Planet formation
models need to robustly incorporate geochemistry/volatiles/isotopic signatures to consider the hypothetical
habitability of exoplanets formed under very different system architectures than our own. Armed with our
understanding of the divergent habitability paths of Earth, Venus, and Mars, we can look outward to the array
of terrestrial exoplanets to be discovered to understand their potential to host and sustain life.


Izidoro, A., A. Morbidelli, and S. N. Raymond (2013), Terrestrial planet formation in the presence of migrating super-Earths,


Stevenson, D. J. (2003), Styles of mantle convection and their influence on planetary evolution, Comptes Rendues de L’Academie des Sciences, 335, 99–111.


