Measuring heritability in messy prebiotic chemical systems

Tymofii Sokolskyi[1,2], Sydney Gargulak[1], Esau Allen[1], David Baum[1,2]
1 - Wisconsin Institute for Discovery, University of Wisconsin-Madison, Madison, WI 53715, USA
2 - Department of Botany, University of Wisconsin-Madison, Madison, WI 53706, USA

Abstract

A key question in origins-of-life research, is whether heritability, and thus evolution, could have preceded genes. While theoretical research has shown that out of equilibrium chemical reaction networks with multiple autocatalytic motifs can provide chemical "memory" and serve as units of heritability, experimental validation is lacking. We established conditions that may be conducive to these processes and developed methods to seek evidence of heritability. We prepared a food set (FS) of three small organic species (methanol, formic acid, and acetic acid), three inorganic salts (sodium trimetaphosphate, ammonium hydroxide, sodium bicarbonate), and pyrite mineral. We conducted a serial dilution experiment where FS was assembled, autoclaved, and the incubated for 24 hours, after which a 20% fraction was transferred into freshly prepared FS that went through the same procedure. This process was repeated for 10 transfer (TR) generations. To serve as controls, we also incubated the fresh solutions in each generation. Over the course of the experiment, we compared the chemical composition of transfer (TR) vials and no-transfer control (NTC) vials using liquid chromatography-mass spectrometry (LCMS). We adapted metrics from ecology and evolutionary biology to analyze the chemical composition data, finding a great deal of variability. Nonetheless, by focusing on a subset of the chemicals with more consistent patterns, we found evidence of heritable variation among vials.

Introduction

The capacity for evolution is one of the key attributes of life (Joyce, 1994). As a result, finding chemical systems that are capable of adaptive evolution, yet simple enough to emerge spontaneously under prebiotic conditions is a central goal of origins-of-life research (Paul & Joyce, 2004). A necessary prerequisite for adaptive evolution is a mechanism to generate heritable variation. In modern evolutionary biology, heritability is due to the presence of variation in self-replicating genetic molecules, which tends to result in a correlation between the traits of parents and their progeny. However, theoretical work suggests that heritable variation need not depend upon the existence of genetic molecules like RNA or DNA but can also arise in chemical reaction networks (CRNs) that contain self-amplifying, which is to say autocatalytic, motifs (Wachtershauser 1988; Segre et al., 2001a; Vasas et al. 2012; Baum 2018). However, to date, experimental data validating these claims are lacking.

It has been proposed that life emerged when CRNs driven out of equilibrium by a continuing influx of "food" established some form of spatial structure, for example in micelles (Segre et al., 2001a) or mineral surfaces (Wachtershauser et al. 1988; Baum, 2018). Given occasional seeding of new autocatalytic motifs by rare seeding events, different spatial locations can establish different dynamical states that are then passed on to other locations via dispersal (Peng et al., 2022; Baum et al. 2023). However, such autocatalytically "encoded" heritability has yet to be shown in the lab and, indeed, evidence of autocatalysis in small-molecules organic chemistry remains scant (Orgel, 2008).

Since heritability depends on their being multiple autocatalytic motifs (Vasas et al., 2015), a minimal requirement for a CRN to show evolvability is it generate many different chemical compounds. Such compositional diversity is readily achieved in prebiotic chemistry via combinatorial explosions (Warr, 1997; Colon-Santos et al., 2019; Matange et al., 2023), especially in the presence of heat and catalytic surfaces (de Graaf et al., 2023). It seems theoretically possible, therefore, to initiate combinatorial explosions in the lab in such a way as to foster sufficient autocatalytic complexity to result in the emergence of heritable variation.

Despite the desirability of creating laboratory conditions suitable for the emergence of heritability and evolution, there remain many challenges. These first is to establish independent systems that can be interpreted as "parents" and "offspring," which can be solved by artificially imposing parent-offspring kinship and analyzing multiple independent lineages that are driven from chemical equilibrium by the same chemical input (Baum & Vetsigian, 2017). This could be achieved using a set of continuously stirred tank reactors (Happel & Stadler, 1999) or via recursive transfer experiments, also known as chemical ecosystem selection (Vincent et al., 2019).

Assuming one has multiple parent-offspring lineages that have been chemically characterized, an obvious approach to measuring heritability, analogous to methods used in plant and animal breeding, would be to calculate the slope of a regression of parent vs. offspring trait values (Sokolskyi et al., 2024). More sophisticated methods that exploit multivariate data are desirable, however, since they may be better able to separate signal from environmental noise. One such an approach was developed by Guttenberg et al. (2015), which uses principal component analysis to estimate the number of heritable states in a complex system. Another approach was proposed by Segrè et al. (2001b) based on the similarity among vectors of chemical composition.

In this paper we used a chemical ecosystem selection paradigm in conditions where a simple, small-molecule food solution yields a combinatorial explosion of chemical complexity. Then, using the detailed chemical fingerprints generated by liquid-chromatography/mass-spectrometry (LCMS), we applied various heritability metrics to compare among multiple parent-offspring lineages. We find that the chemical composition of systems is sensitive to a history or recursive transfers and that "offspring" systems tend to be more similar to their "parents" than their non-parents. This analysis suggests promising avenues for further research and challenges that should be born in mind when designing future prebiotic chemistry experiments that search for the *de novo* emergence of heritability and evolutionary change.

Materials & Methods

Compound	Concentration	Vendor	Part number
Formic acid	10 mM	Aqua Solutions	F6088
Acetic acid	10 mM	Fischer Scientific	BP2401
Methanol	10 mM	Thermo Fischer Scientific	A4564
Ammonium hydroxide	4 mM	VWR	EM1.05432.1011
Sodium bicarbonate	20 mM	Sigma Aldrich	S6014

Sodium trimetaphosphate	0.1 mM	Sigma Aldrich	T5508
Pyrite	100 mg/mL	VWR	100198-728

Table 1. Composition of the food set (FS).

Solution preparation. We prepared a food set (FS) solution consisting of 6 compounds and a mineral, pyrite (**Table 1**), which is known for its catalytic activity (de Graaf et al., 2023). All components except pyrite were mixed in water and stored at 4°C until use. An aliquot of FS was added to pyrite every generation of the experiment. The organic components, formic acid, acetic acid and methanol, are prebiotically plausible, and were likely abundant on early Earth (McColom & Seewald, 2007; Mohammadi et al., 2020). Bicarbonate may support pyrite-catalyzed carbon fixation, especially under high temperature and pressure (Cody et al., 2004). Ammonium hydroxide and trimetaphosphate serve as sources of nitrogen and phosphorus respectively.

Experimental design. We conducted chemical ecosystem selection involving a transfer-with-dilution protocol (**Fig. 1**). First, we aliquoted 2 mL of FS with pyrite into glass vials (Neta Scientific, RES-24658), sealed with rubber stoppers (Neta Scientific, W224100-400) and aluminum seals (Neta Scientific, 224177-05). This procedure was conducted in an anaerobic chamber (95% N2, 5% H2). The starting vials, designated generation 0, were autoclaved using a liquid cycle (30 mins, 121°C, 30 psi) and incubated for 24 hours in a 40°C orbital shaking incubator. Generation 1 vials received 1.6 mL FS and ~200 mg pyrite and either 0.4 mL (20%) of the solution from a "parental" generation 0 vial or an additional 0.4 mL of fresh FS. Vials receiving transfers (TRs) and no transfer controls (NTCs) were again autoclaved and incubated for 24 hours. This procedure was repeated for 10 generations, with vials from each generation stored at -80°C until analysis. In this work, our main analytical focus was to compare TRs to NTCs every generation, as it allows us to control for experimental and analytical artifacts. We conducted two identical replicate experiments, each with 10 TR lineages and 10 NTC pseudo-lineages per generation. The results shown in this study are based on a combined analysis of both experiments.

To help interpret the results and explain the role of pyrite and autoclaving in the combinatorial explosion, we examined changes in chemical composition in FS over a single 24-hour incubation period. FS solution was autoclaved and incubated or simply incubated with and without pyrite, with samples taken every 2 hours. In addition to FS, we conducted a similar procedure for samples that lacked subsets of FS components.

Liquid-chromatography/mass-spectrometry (LCMS). Immediately prior to chemical analysis, vial contents were thawed and transferred to 96-well 0.2 mm filter plates (Neta Scientific, PALL-8019) and vacuum filtered, to remove pyrite particles. We used an untargeted metabolomics approach with reverse phase Ultra-Performance Liquid Chromatography coupled with tandem mass-spectrometry (UPLC-MS/MS) with a Thermo Vanquish UPLC (Thermo Fisher Scientific, Waltham, USA) with a C18 column (Agilent, Santa Clara, USA). Samples (5 µL) were eluted in a linear gradient mixture from 0.1% v/v formic acid in water (47146-M6, Thermo Fisher Scientific, Waltham, USA) to 0.1% v/v formic acid in acetonitrile (47138-K7, Thermo Fisher Scientific,

Waltham, USA), over 14 min. To reduce the risk that order effects could yield artifactual differences between TR and NTC treatments, samples were run with the first TR and NTC replicates preceding the second replicate of each, and so on.

The UPLC was coupled to a Thermo Q-Exactive Plus Orbitrap MS (Thermo Fisher Scientific, Waltham, USA). Full MS-SIM spectra for each of the replicates were collected for 10 minutes in positive mode over a scan range of 50–750 m/z, with resolution set to 70000. Fragmentation data were collected only for pooled samples for each treatment category with full MS followed by data-driven MS2 analysis (dynamic exclusion of 4s, intensity threshold 1.0E5, resolution 17500, isolation window 1 m/z, stepped collision energies of 20, 50 and 100 eV).

Data structure. Data in .raw format was processed in Compound Discoverer™ (Thermo Fisher Scientific, Waltham, USA). We used a default untargeted metabolomics workflow with 16 added databases (Across Organics, Alfa Aesar, Alfa Chemistry, BioCyc, Cambridge Structural Database, CAS Common Chemistry, ChemBank, DrugBank, FDA, Human Metabolome Database, KEGG, MassBank, Merck Millipore, MeSH, NIST, NPAtlas) with pooled samples set to 'identification only'. Every sample from each experiment was processed in the same Compound Discoverer analysis. Lists of identified compounds, their relative abundances and SMILES IDs were extracted for downstream analyses. Additionally, for samples used to evaluate the role of autoclaving and pyrite in generating chemical complexity, we ran samples individually in Compound Discoverer™ in order to obtain comparable counts of the number of MS features. Lists of identified compounds in the combined Compound Discoverer™ analysis, their relative abundances and SMILES IDs were extracted for downstream analyses.

All features without assigned names or formulas were removed from the data due to uncertainty about their identity and consequent difficulty in ensuring comparability between samples. Areas of MS features were summed if they had the same name or, if names were absent, the same formula. The resulting dataframes are lists of areas of identified compounds in each sample and their assigned names/formulas.

To evaluate whether the TR and NTC samples in a generation are different, and whether the degree of difference changes over generation, we calculated Canberra (**Eq. 1**) and Bray-Curtis dissimilarities (**Eq. 2**) between each pair of samples using the vegan package in R (Oksanen et al., 2007). These distance metrics are frequently used to quantify dissimilarity between ecological communities (Ricotta & Podani, 2017). Since the Canberra distance sums ratios rather than areas, it tends to weight each species more or less equally, regardless of its absolute abundance, unlike Bray-Curtis which assigns higher weight to more abundant species.

$$d_{p,q} = \frac{1}{n}\sum_{i+1}^{n}\frac{|p_i - q_i|}{|p_i| + |q_i|}$$

Equation 1. Formula for Canberra distance: *n* – number of compounds; *i* – individual compound; *p* – area of compound *i* in one sample; *q* – area of a compound *i* in a different sample.

$$d_{p,q} = \frac{\sum_{i=1}^{n} |p_i - q_i|}{\sum_{i=1}^{n} (p_i + q_i)}$$

Equation 2. Formula for Bray-Curtis distance: *n* – number of compounds; *i* – individual compound; *p* – area of compound *i* in one individual; *q* – area of a compound *i* in a different individual.

To evaluate whether TR and NTCs differ, we compared the pairwise distances between samples within groups (TRs and NTCs) with the between group distance (NTC-TR). We conducted a permutational analysis of variance (PERMANOVA) in R. We calculated the mean sum of square Canberra and Bray-Curtis distances and obtained standard deviations based on 500 bootstrap resamplings of features.

Heritability calculations. In evolutionary biology heritability is typically estimated as the slope of a linear regression of trait values that have been measured in a set of parents and their corresponding offspring. Since we have 20 parent/offspring pairs in each generation, we have the capacity to calculate an analog of heritability for these data.

The first method we used is similar to the biological method and involves calculating the Pearson correlation of the area of a compound in parents and offspring in adjacent generations (e.g., generations 1 and 2, 2 and 3, etc.). We then compared the regression slope, r, over all compounds to see if the null hypothesis of r=0 can be rejected using a t-test. We also calculated the regression-based correlation for two summary statistics, namely the total number of significantly enriched or depleted compounds relative to the set of NTCs from that generation.

To supplement the biological approach, we also developed a multivariate measure of heritability, the Distance-base Heritability Index (DHI; **Eq. 3**). This index divides the distance between parent-offspring pairs in adjacent generations with the mean distance between non-parent-offspring pairs. Observing DHI values that are lower for TRs than NTCs would show that samples from the same lineage are more similar to each other than would otherwise be expected, thus supporting heritability. Our DHI calculations used the Canberra distance, which weights all compounds equally, regardless of mean concentration.

$$DHI^{g,l,x} = L(L-1)\frac{CD_{g+1,l,x}^{g,l,x}}{\sum_{l' \neq l''} CD_{g+1,l',x}^{g,l'',x}}$$

Equation 3. *DHI* – distance-based heritability index; *CD* – Canberra distance; *g* – generation number; *l* – lineage number; *x* – sample type (TR or NTC); *L* – total number of lineages.

Finally, we applied a heredity metric developed by Guttenberg et al. (2015) for use in studies of prebiotic chemistry. We used their Python script (https://github.com/ModelingOriginsofLife/Heredity/tree/master) to analyze our data and estimate the number of heritable states possible for the TRs, as compared to the NTCs. This method conducts principal component analysis on an input set of properties for systems that

are inferred to have been exposed to different selective environments. The number of eigenvalues above a computed threshold is taken as an estimate of the number of potentially heritable states. This methodology can distinguish between intrinsic (due to stochastic processes) and extrinsic (due to selection) sources of variation because, intrinsic variation should be unimodal and not dependent on changes to the fitness landscape, while extrinsic variation should be multimodal, around distinct fitness optima (Guttenberg et al., 2015). We assumed that each generation experienced a somewhat different selective pressure, as measured by the amount of generation 0 contents transferred, and estimated the number of heritable states per experimental lineage (N_S) with the prediction of lower N_S in NTCs, which lack inter-generational transmission. Prior to this calculation, we normalized our dataset by subtracting the mean and dividing by the standard deviation for each compound, as is a common practice for principal component analysis (Guttenberg et al., 2015). The parameter alpha, which sets the sensitivity of the algorithm, was set to 0.1.

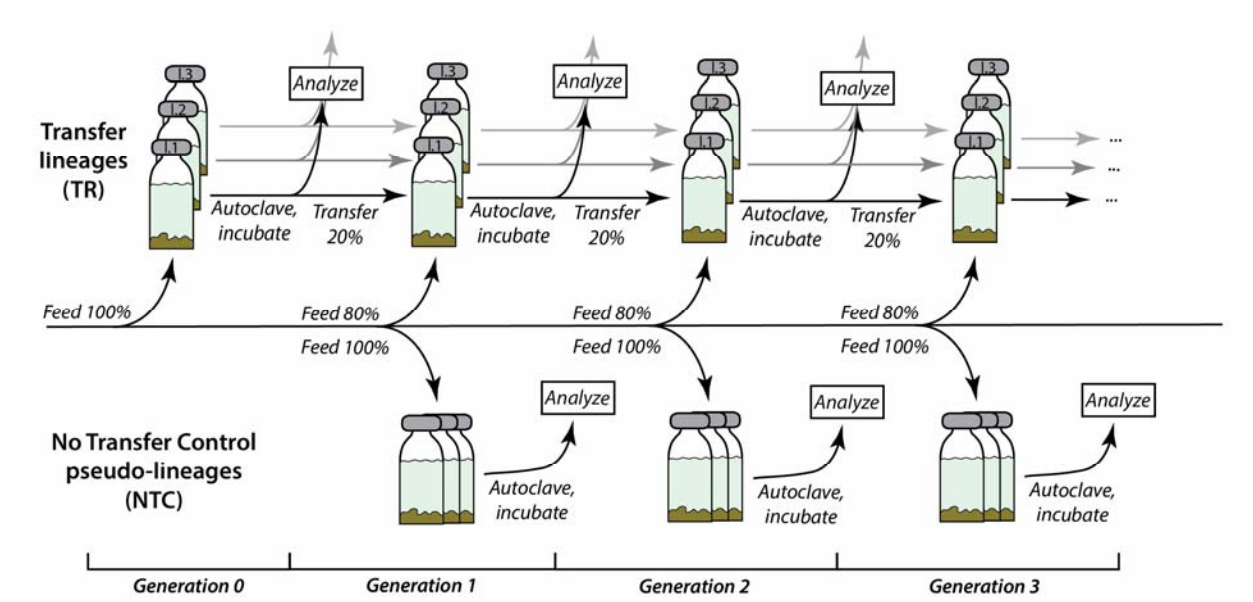


Figure 1. Scheme of the experimental design.

Results

Pyrite and autoclaving promote chemical diversity. To test for the existence of heritability, it was necessary that the experimental system generated sufficient chemical complexity that different lineages could, in principle, diverge from one another. Based on individual Compound Discoverer™ runs we evaluated the average number of features above an arbitrary area threshold of 10^6 for different treatments (**Supp. Table B1**). The largest factor seemed to be the presence/absence of pyrite since more than 500 features were detected in samples with FS and pyrite, compared to 166 without. The massive increase in the number of compounds seen after incubation with pyrite with or without autoclaving can plausibly be explained by the chemistry of the food as summarized in **Appendix A**. However, almost 500 features were also detected in samples with water + pyrite, which might suggest that the effect of pyrite is due to contamination. While we cannot rule this out, it is noteworthy that, when analyzed together,

most compounds found in water + pyrite are undetected or less abundant in FS + pyrite (**Fig. 2**). As result, it is possible many represent analytical artifacts rather than contamination from pyrite. Sodium and potassium salt clusters and other metal ions are a common electrospray ionization (ESI) artifact (McMillan et al., 2016), and likely represent a significant amount of detected features, since FS includes two sodium salts (bicarbonate and trimetaphosphate). Differences in pyrite oxidation between water and FS could also affect feature counts as Fe adducts are another ESI artifact (Bonner & Hopfgartner, 2022). Hence it is unclear whether absolute feature counts accurately represent chemical diversity of our samples.

Combined analysis in CompoundDiscoverer™ makes it possible to directly compare the mass spectral features, at the cost that all compounds that are detected in any sample are assigned an imputed area in all other samples, whether or not they were detected. To get around this problem and quantitatively assess how pyrite and autoclaving affect compound abundances, we counted up the number of compounds that have significantly elevated areas in one treatment relative another (**Fig. 2)**. The effect of autoclaving was assessed by comparing FS without pyrite to autoclaved FS. The effect of pyrite was determined by comparing unautoclaved FS with pyrite to unautoclaved FS without pyrite. These analyses revealed that an "explosion" of chemical diversity occurred both due to autoclaving and presence of pyrite, with pyrite having a greater and more immediate effect.

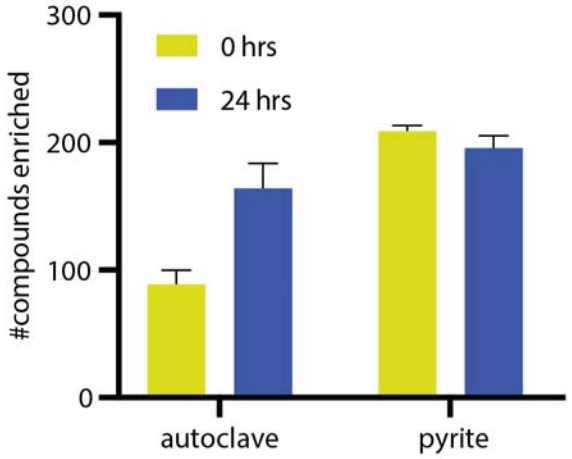


Figure 2. Number of detected compounds with area >10^6 significantly enriched relative to respective controls (>2*SD difference: compared to unautoclaved samples without pyrite for autoclave effect; compared to unautoclaved samples without pyrite and to water with pyrite for pyrite effect) in FS due to autoclaving or pyrite. Error bars are SEM.

Transfer samples diverge from controls. To quantify the divergence between transfer lineages and controls in the multi-generation experiments, we conducted PERMANOVA analysis on Canberra and Bray-Curtis distances. For Canberra distances NTC-TR distances are greater than either NTC-NTC or TR-TR distances, except for generation 6 when TR-TR distances are higher. This difference was significant ($p<0.05$) in every generation except generation 5 (**Fig. 3, A**). Bray-Curtis distances also tend to be more similar within than between groups, but NTC-TR and NTC-

NTC are significant different only in generations 3, 4 and 6 (**Fig. 3, B**). Taken together, the tendency for greater NTC-TR distances than within-group distances shows that the presence of material transferred from prior generations significantly affects the overall composition of vials.

In addition to distance metrics, we also calculated the number of compounds that are significantly (>2*SD_{NTC} difference) enriched and depleted in TRs compared to NTCs from their generation (**Fig. 3, C**). Since TRs carry forward products from one generation to the next, we predicted that there will often be compounds that are only seen above the threshold of detection in TRs. Indeed, about ten times as many compounds are enriched than depleted when comparing TRs and NTCs. While the number of depleted compounds remains similar across all 10 generations, the fraction of compounds that are significantly enriched in TRs varies markedly across generations (9-22%).

We calculated the Pearson correlation between the number of enriched and depleted compounds in parent-offspring pairs in adjacent generations (**Fig. 3, D**). While the slope of the regression of depleted compounds randomly fluctuates, there is generally a positive correlation for enriched compounds with a significant slope ($p<0.05$), which can be seen as suggesting heritability for five generation pairs for enriched compounds but only one for depleted compounds.

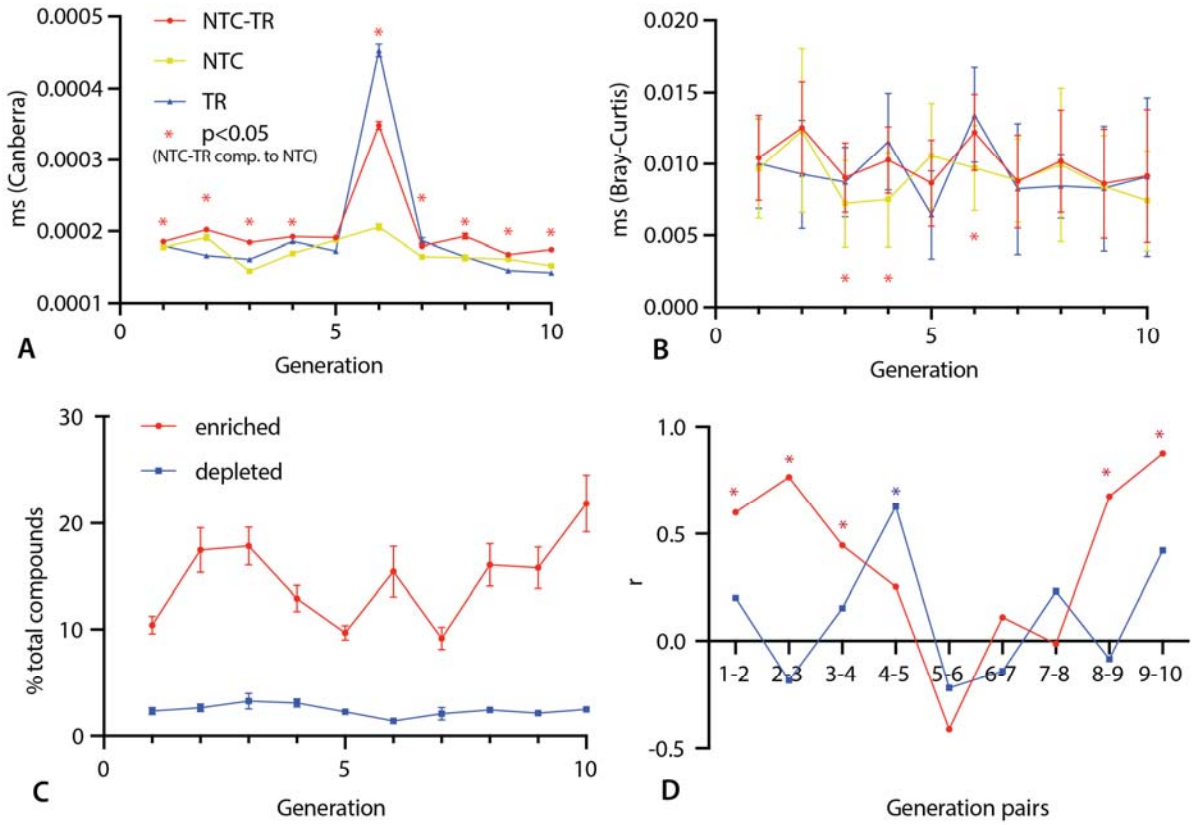


Fig. 3. Dynamic changes over the course of the experiment. **A** – mean sum of squares (ms) of inter-group Canberra distance, plotted are TR-NTC, NTC-NTC and TR-TR group distances; **B** – mean of squares (ms) of inter-group Bray-Curtis distance, plotted are TR-NTC, NTC-NTC and TR-TR group distances; **C** – change in the number of compounds significantly (difference greater

than 2* SD_{NTC}) enriched or depleted in TRs compared to NTCs every generation; **D** – Pearson correlation of the number of enriched and depleted compounds between adjacent generations in the TR treatment (significant correlation, $p<0.05$, is marked by asterisks). Error bars are SD in **A-B** and SEM in **C**.

Heritability measurements. Many individual compounds have more significantly positive parent-offspring correlations than expected by chance. However, a similar fraction was seen in both TRs and NTCs despite the fact that NTCs have no true "parent-offspring" relationships because pairs are simply the same numbered replicate in each generation. Similarly, although average regression slopes are significantly greater than 0 in many generations, these cases were found in both TRs and NTCs with no clear pattern (**Fig. B1**). The most likely explanation of these patterns is not heritability but order of analysis, reflecting the fact that lower-numbered replicates tended to be loaded earlier in an LC-MS run than high-numbered replicates.

In order to obtain a single metric that summarizes all data, we developed the DHI statistic, which measures whether parent-offspring pairs are more similar than unrelated vials. When calculating DHI using the Canberra distance matrix for the complete, filtered dataset we observed no difference between TR and NTC values, which are all in 90-100% range, except for generation 1, where the DHI of TRs is significantly ($p<0.05$) greater than in NTCs (**Fig. 4, A**).

Given the possibility that a strong order-of-analysis effects might be swamping true signals of heritability we recalculated DHI based only on compounds that are the strong candidates for being heritable in TRs but not NTCs and, as a control, those that are candidates for showing higher "heritability" in NTCs. To be a candidate in either direction, a compound showed Pearson correlation coefficients greater than 0.5 in 3 or more pairs of generations for one of the data sets (TR or NTC), but for 1 or fewer pairs of generations for the other data set. Recalculating DHI for the two heritable candidate sets yielded different results for TR and NTC data sets (**Fig. 4, B-C**). In both cases the DHI for the candidates is lower than in either the complete data or for the same compounds in the other data set. However, the DHI decline was greater for the TR heritable core (**Fig. 4B)** than for the NTC control core (**Fig. 4C)**. Statistical tests never detected a significant difference in DHI between NTCs and TRs for the NTC candidates, but the TR DHI is significantly lower than those for NTCs in 6 out of 9 generation pairs for the TR heritable core (**Fig. 4, B-C**).

As a final assessment we also estimated the number of possible heritable states for the full dataset and two candidate subsets, using the method of Guttenberg et al. (2015). For the total dataset we find that TR lineages have more possible heritable states with a close to significant p-value of 0.06 (**Fig. 5**). For the TR candidate core the difference is even more significant ($p = 0.02$), while for the NTC candidate no significant difference was detected ($p=0.37$).

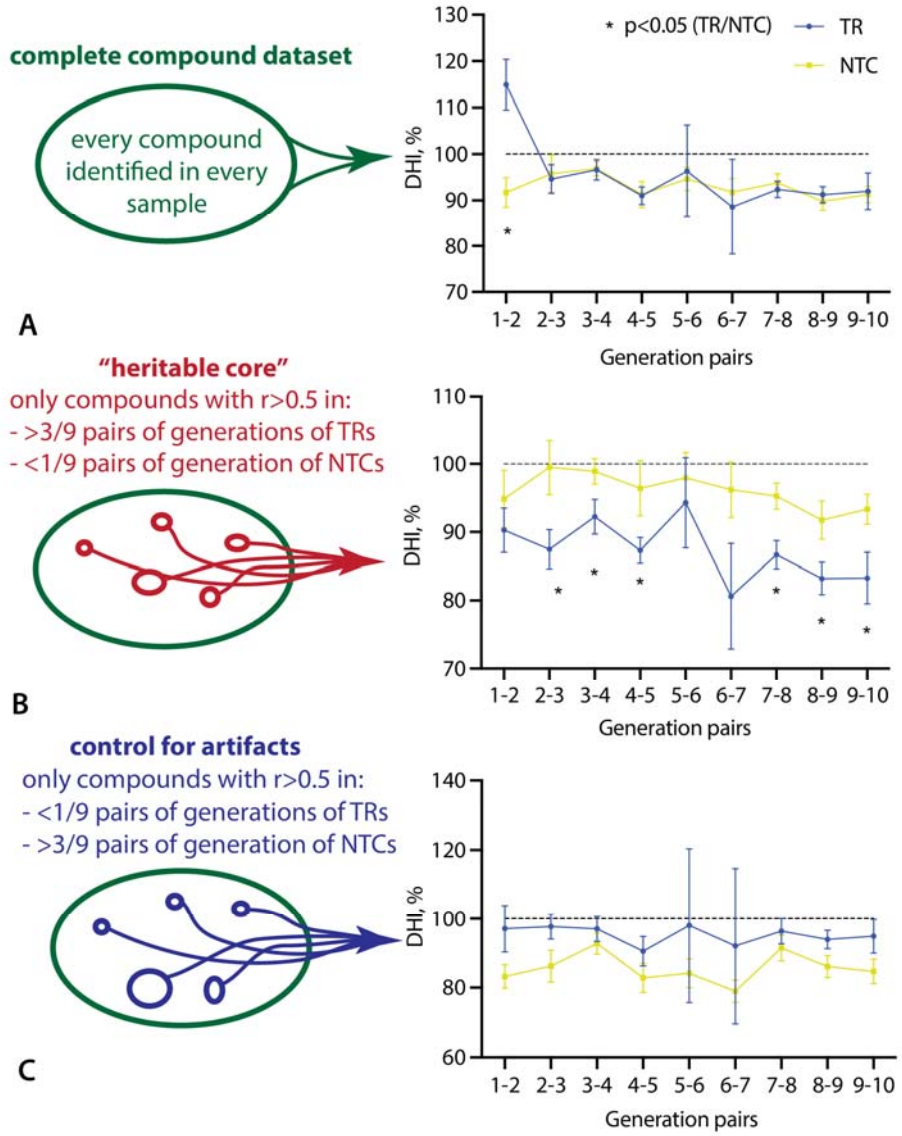


Fig. 4. Changes in heritability as measured by the distance-based heritability index (DHI) using Canberra distance matrix. **A** – change in DHI over the course of the experiment for the whole compound dataset; **B** – change in DHI for the dataset comprising the "heritable core" – compounds that have r>0.5 in 3 or more pairs of generations of TRs, and 1 or less generations of NTCs; **C** – change in DHI for the control dataset – compounds that have r>0.5 in 1 or less pairs of generations of TRs, and 3 or more generations of NTCs. Error bars are SEM.

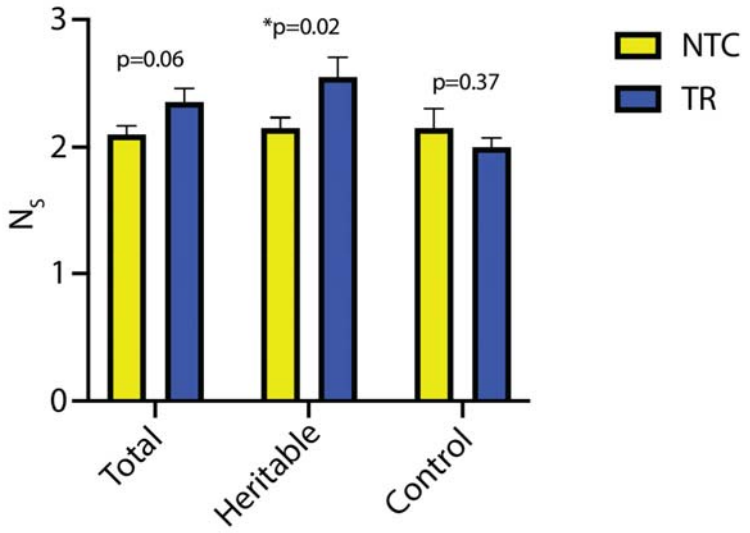


Fig. 5. Number of heritable states (N_S), for the total dataset, heritable core (only compounds with r>0.05 for >3 pairs of generations of TRs and <1 of NTCs) and control core (only compounds with r>0.05 for >3 pairs of generations of NTCs and <1 of TRs). Error bars are SEM.

Discussion

Autoclaving, pyrite and transfers significantly impact chemical diversity. Both autoclaving and pyrite significantly result in the appearance of many more than the original handful of compounds in FS (**Fig. 2**). Pyrite had the greatest effect, presumably because of the ability of pyrite and other metal-rich minerals to catalyze prebiotically-relevant reactions, including abiotic carbon fixation, incorporation of nitrogen into organic compounds, and the formation of larger organics (de Graaf et al, 2023; Asche et al., 2024). Carbon fixation with CO_2, or similar reactions starting with 1 and 2 carbon organic molecules, seem particularly relevant and necessary to explain the many larger organic compounds we detected after incubation. Drawing on changes in vial composition over the course of a 24-hour incubation, we can speculate regarding a few chemical reactions that might be occurring during incubation with pyrite (**Appendix A**).

In order to have any possibility of revealing heritability, it is first essential that the transfer process affects the chemical composition of TR vials relative to NTC vials, which receive no inputs from the prior generation. Indeed, we find clear evidence of a systematic differences between TR and NTC vials in almost all generations, as shown by greater average TR-NTC Canberra distances than either TR-TR or NTC-NTC (**Fig. 3, A**). The fact that this effect is less clear using Bray-Curtis distances (**Fig. 3, B**) likely reflects the fact that this distance measure tends to be dominated by a few high-abundance compounds.

Heritability can be detected in small-molecule chemical systems. In this study we used a structure with multiple parallel parent-offspring transfer lineages to detect heritable variation emerging from prebiotically plausible reactions. Here we understand heritability, analogously to biology, as the fraction of the variation between individuals in a population that is inherited

354 from parents (Visscher et al., 2008). In this context, the 20 TR vials in adjacent generations
355 comprise the population of parents and offspring and the phenotypes are areas of different
356 mass-spectral features.
357 To be relevant to long term evolution, heritable differences among vials would need to
358 be the result of autocatalytic cycles that are activated in some but not all vials. However, any
359 given autocatalytic cycle may have only a small number of detectable members meaning that
360 only a tiny fraction of compounds will show heritability, with the remaining variation being due
361 to experimental noise. Thus, the fact that we did not find many more compounds in TR vials
362 that had high parent-offspring correlations is insufficient to reject the hypothesis of heritability.
363 Three analyses supported the idea that heritability is present in this system. First, two
364 system-level, emergent traits, the number of enriched and depleted compounds, show patterns
365 consistent with heritability (**Fig. 3, D**). Second, we found DHI is significantly higher for the TR
366 heritable candidates in most generations, whereas no such pattern in seen using the NTC
367 candidates (**Fig. 4**). Third, the number of heritable states in TRs tends to be significantly greater
368 than in NTCs (**Fig. 5**). Collectively, these metrics provide the first direct evidence that chemical
369 systems driven from equilibrium by prebiotically plausible chemical food sets can vary and pass
370 on those variations into the future. This is important because the generation of heritable
371 variation is a minimal requirement for evolution by natural selection.
372 To evaluate the significance of this result, it is worth considering whether these patterns
373 could arise by trivial mechanisms, of which the most obvious is contaminants being introduced
374 into a subset of parent vials that are then seen at elevated concentrations in offspring vials
375 simply because the latter receive, as input, 20% of the volume from the parent vial. If such
376 contaminants did not react, we would expect them to occur at five-fold lower concentration in
377 offspring than parents. A positive correlation between parents and offspring could be seen,
378 nonetheless, but only if the magnitude of the contamination received by parents is five times
379 greater than the threshold of detection by LC-MS. We believe that such high-concentration
380 contaminants are likely rare, so we take the evidence of heritability as suggestive of
381 autocatalysis. Our hypothesis is that variation in parent vials arises from time to time, whether
382 from sub-threshold contamination or stochastic chemical dynamics, and this triggers novel
383 autocatalytic processes that are passed on to offspring vials. Ultimately, to determine if this is
384 the case, it would be necessary to reconstruct the underlying CRN in sufficient detail to know if
385 autocatalytic cycles exist that could explain the observed data, especially the subset of
386 individual compounds that have significant parent-offspring correlations. This would be
387 challenging but may be possible using reaction-rule based computational chemistry, and is
388 certainly worth exploring in the future.
389 Although there is an obvious need for validation, this work provides the first statistical
390 evidence for the spontaneous emergence of heritable variation in prebiotically plausible
391 chemical mixtures. This represents a significant advance towards the goal of understanding
392 how systems capable of adaptive evolution could have bootstrapped themselves into existence
393 on the early Earth. Moreover, while there are areas where the protocol could be improved,
394 most obviously by randomizing LC-MS analysis order to remove order-of-analysis effects in
395 NTCs, our experimental approach and the metrics developed may serve as a blueprint for
396 further investigations into the emergence of heritability and evolvability in the absence of
397 genetic polymers.

Acknowledgments. The authors would like to thank Zoe Todd, John Yin, Betül Kaçar, Daniel Amador-Noguez and Annie Bauer for feedback on this project. We would like to thank Chris Thomas, Julia Duncan and Eric Armstrong for valuable guidance on LCMS operation and data analysis. We would like to thank Praful Gagrani for clarifications on mathematical concepts and Lena Vincent for help with experimental design. We thank Cecile Ane from the Statistical Consulting group at the University of Wisconsin – Madison for assistance in the analysis of the experiment. This research was funded by the Department of Botany at the University of Wisconsin–Madison and the National Science Foundation (grant number DEB 2218817).

References

Asche, S., Pow, R. W., Mehr, H. M., Cooper, G. J., Sharma, A., & Cronin, L. (2024). Evidence of Selection in Mineral Mediated Polymerization Reactions Executed in a Robotic Chemputer System. ChemSystemsChem, 6(3), e202400006.

Baum, D. A., & Vetsigian, K. (2017). An experimental framework for generating evolvable chemical systems in the laboratory. Origins of Life and Evolution of Biospheres, 47, 481-497.

Baum, D. A. (2018). The origin and early evolution of life in chemical composition space. Journal of theoretical biology, 456, 295-304.

Baum, D. A., Peng, Z., Dolson, E., Smith, E., Plum, A. M., & Gagrani, P. (2023). The ecology–evolution continuum and the origin of life. Journal of the Royal Society Interface, 20(208), 20230346.

Bonner, R., & Hopfgartner, G. (2022). Annotation of complex mass spectra by multi-layered analysis. *Analytica Chimica Acta*, *1193*, 339317.

Cody, G. D., Boctor, N. Z., Brandes, J. E. A., Filley, T. R., Hazen, R. M., & Yoder Jr, H. S. (2004). Assaying the catalytic potential of transition metal sulfides for abiotic carbon fixation. Geochimica et Cosmochimica Acta, 68(10), 2185-2196.

Colón-Santos, S., Cooper, G. J., & Cronin, L. (2019). Taming the combinatorial explosion of the formose reaction via recursion within mineral environments. ChemSystemsChem, 1(3), e1900014.

de Graaf, R., De Decker, Y., Sojo, V., & Hudson, R. (2023). Quantifying Catalysis at the Origin of Life. Chemistry–A European Journal, 29(53), e202301447.

Guttenberg, N., Laneuville, M., Ilardo, M., & Aubert-Kato, N. (2015). Transferable measurements of Heredity in models of the Origins of Life. PloS one, 10(10), e0140663.

Guttenberg, N., Virgo, N., Chandru, K., Scharf, C., & Mamajanov, I. (2017). Bulk measurements of messy chemistries are needed for a theory of the origins of life. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 375(2109), 20160347.

Happel, R., & Stadler, P. F. (1999). Autocatalytic replication in a CSTR and constant organization. Journal of mathematical biology, 38(5), 422-434.

Joyce, G. F. (1994). Forward. Origins of Life: The Central Concepts. W. Deamer and GR Fleischaker, eds.

Lu, H., Blokhuis, A., Turk-MacLeod, R., Karuppusamy, J., Franconi, A., Woronoff, G., ... & Griffiths, A. D. (2024). Small-molecule autocatalysis drives compartment growth, competition and reproduction. *Nature Chemistry*, *16*(1), 70-78.

Lush, J. L. (1937) Animal Breeding Plans. Ames, Iowa: Iowa State Press.

Matange, K., Rajaei, V., Capera-Aragonès, P., Costner, J. T., Robertson, A., Kim, J. S., ... & Pinter, M. F. (2023). Evolution of complex chemical mixtures reveals combinatorial compression and population synchronicity.

McCollom, T. M., & Seewald, J. S. (2007). Abiotic synthesis of organic compounds in deep-sea hydrothermal environments. Chemical reviews, 107(2), 382-401.

McMillan, A., Renaud, J. B., Gloor, G. B., Reid, G., & Sumarah, M. W. (2016). Post-acquisition filtering of salt cluster artefacts for LC-MS based human metabolomic studies. *Journal of cheminformatics*, *8*, 1-5.

Mohammadi, E., Petera, L., Saeidfirozeh, H., Knížek, A., Kubelík, P., Dudžák, R., Krůs, M., Juha, L., Civiš, S., Coulon, R. and Malina, O., 2020. Formic acid, a ubiquitous but overlooked component of the early earth atmosphere. Chemistry–A European Journal, 26(52), pp.12075-12080.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. A. S. S. (2007). The vegan package. Community ecology package, 10(631-637), 719.

Paul, N., & Joyce, G. F. (2004). Minimal self-replicating systems. Current opinion in chemical biology, 8(6), 634-639.

Peng, Z., Kaçar, B., Fahrenbach, A., & Adam, Z. (2023). An assessment of stoichiometric autocatalysis across element groups. chemrxiv

Peng, Z., Linderoth, J., & Baum, D. A. (2022). The hierarchical organization of autocatalytic reaction networks and its relevance to the origin of life. PLOS Computational Biology, 18(9), e1010498.

Peng, Z., Plum, A. M., Gagrani, P., & Baum, D. A. (2020). An ecological framework for the analysis of prebiotic chemical reaction networks. Journal of theoretical biology, 507, 110451.

Ricotta, C., & Podani, J. (2017). On some properties of the Bray-Curtis dissimilarity and their ecological meaning. Ecological Complexity, 31, 201-205.

Segré, D., Ben-Eli, D., Deamer, D. W., & Lancet, D. (2001a). The lipid world. Origins of Life and Evolution of the Biosphere, 31, 119-145.

Segré, D., Shenhav, B., Kafri, R., & Lancet, D. (2001b). The molecular roots of compositional inheritance. Journal of Theoretical Biology, 213(3), 481-491.

Sokolskyi, T., Ganju, P., Montgomery-Taylor, R., & Baum, D. A. (2024). Evidence of Heritability in Prebiotically Realistic Membrane-Bound Systems. Life, 14(3), 284.

Srivastava, R., Dueñas-Díez, M., & Pérez-Mercader, J. (2018). Feed rate noise modulates autocatalysis and shapes the oscillations of the Belousov–Zhabotinsky reaction in a continuous stirred tank reactor. Reaction Chemistry & Engineering, 3(2), 216-226.

Vasas, V., Fernando, C., Santos, M., Kauffman, S., & Szathmáry, E. (2012). Evolution before genes. Biology direct, 7, 1-14.

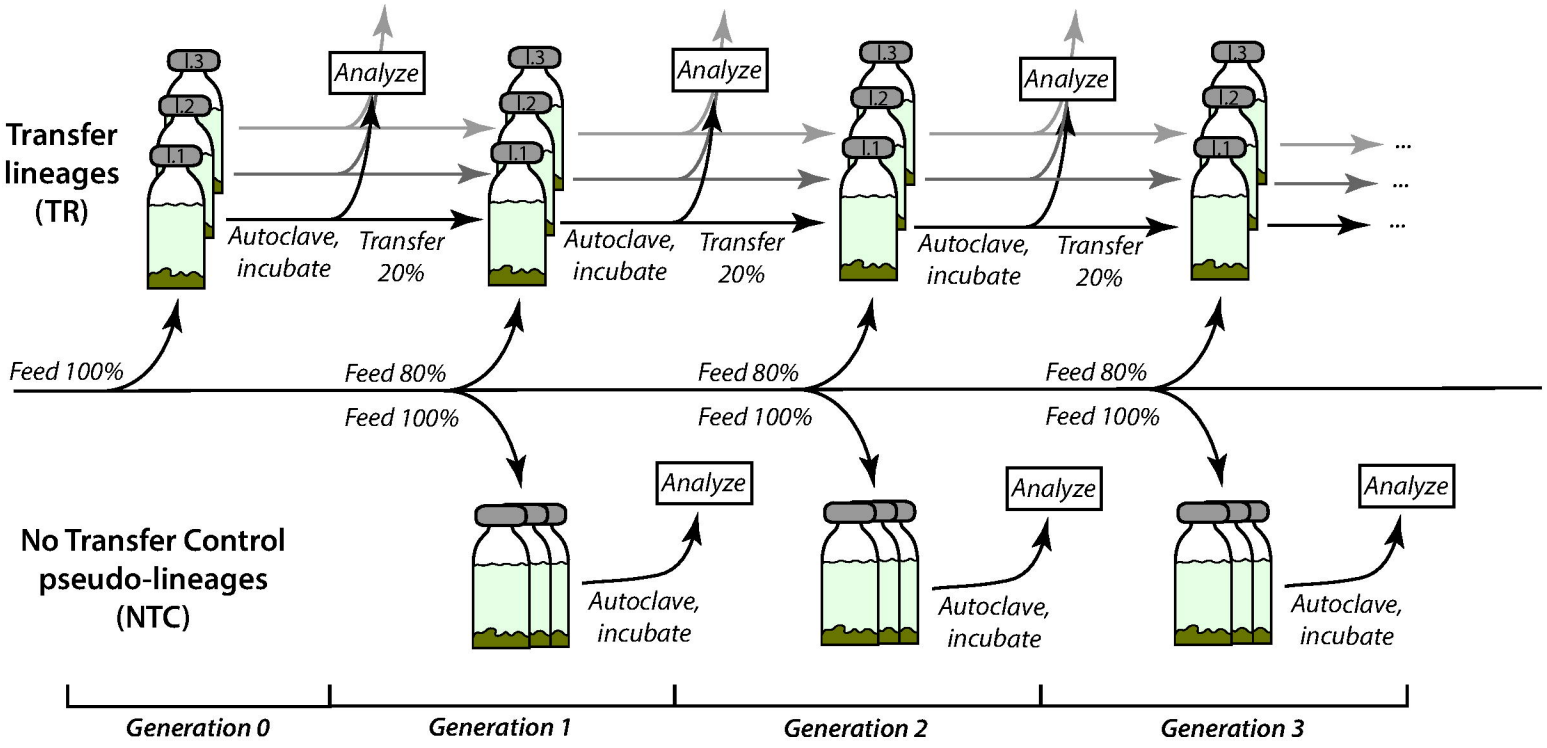
Vasas, V., Fernando, C., Szilágyi, A., Zachár, I., Santos, M., & Szathmáry, E. (2015). Primordial evolvability: Impasses and challenges. Journal of theoretical biology, 381, 29-38.

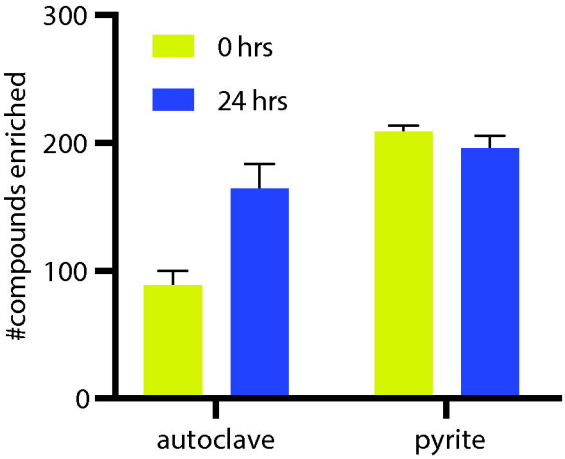
Vasas, V., Szathmáry, E., & Santos, M. (2010). Lack of evolvability in self-sustaining autocatalytic networks constraints metabolism-first scenarios for the origin of life. Proceedings of the National Academy of Sciences, 107(4), 1470-1475.

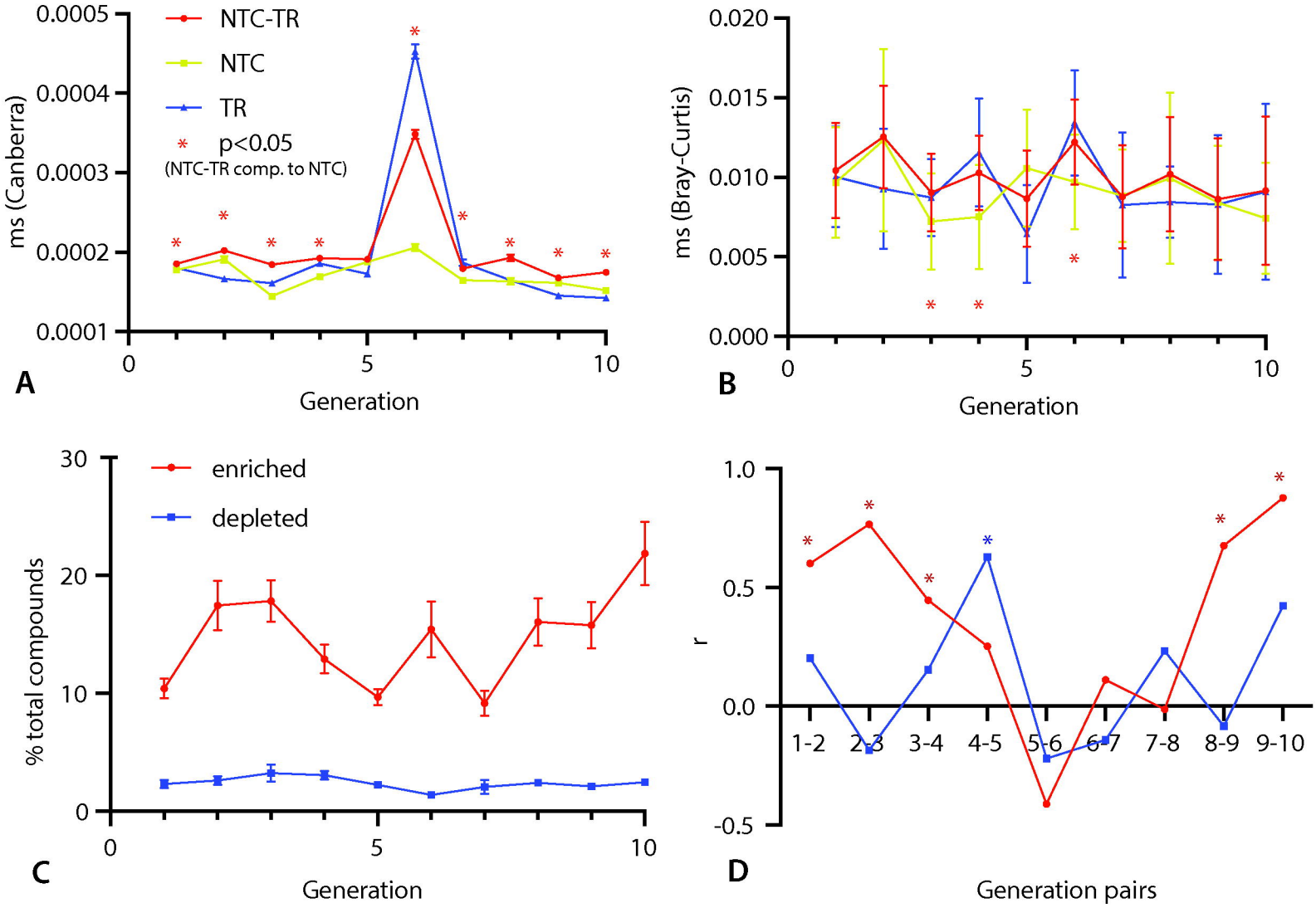
Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era—concepts and misconceptions. Nature reviews genetics, 9(4), 255-266.

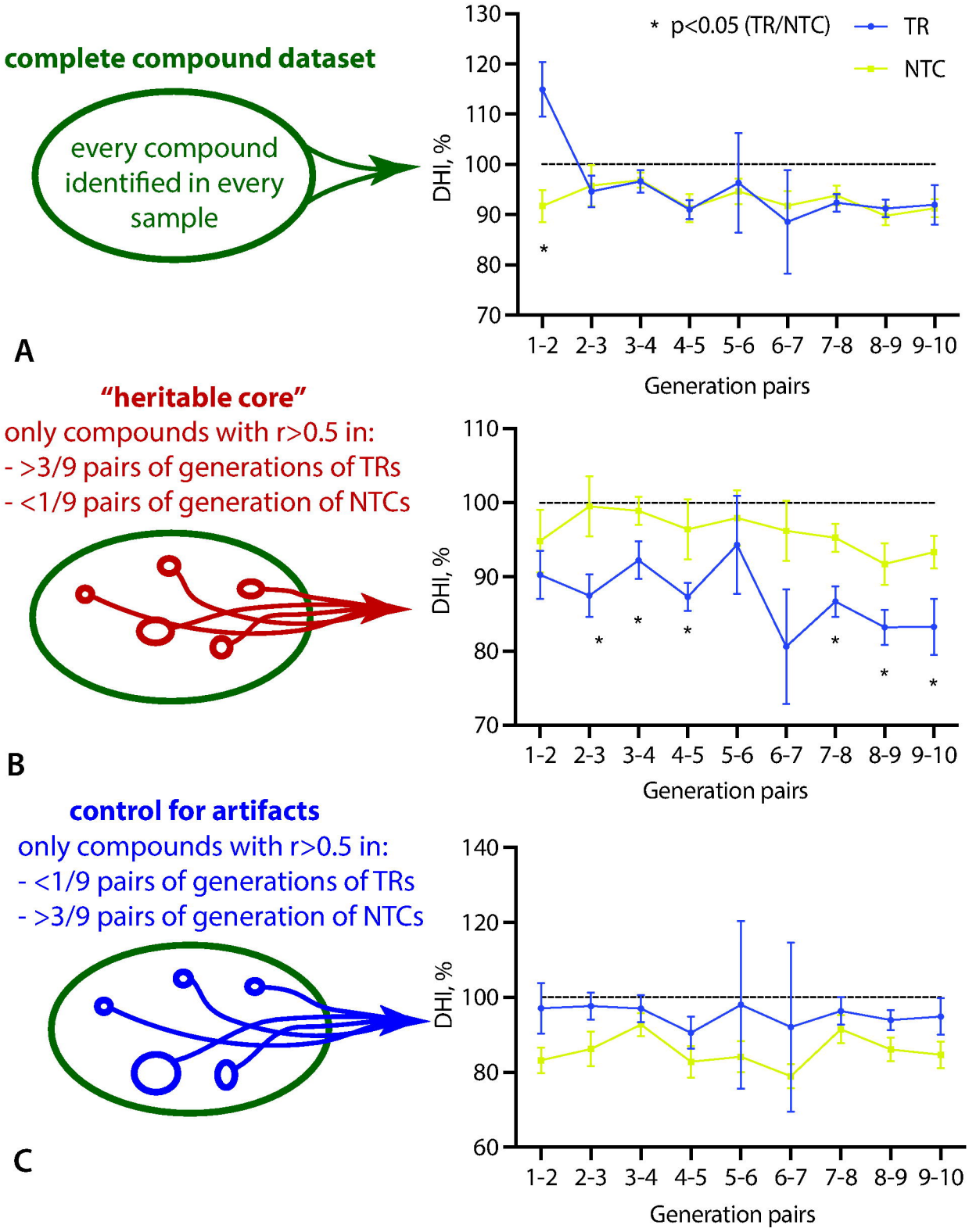
Wächtershäuser, G. (1988). Pyrite formation, the first energy source for life: a hypothesis. Systematic and Applied Microbiology, 10(3), 207-210.

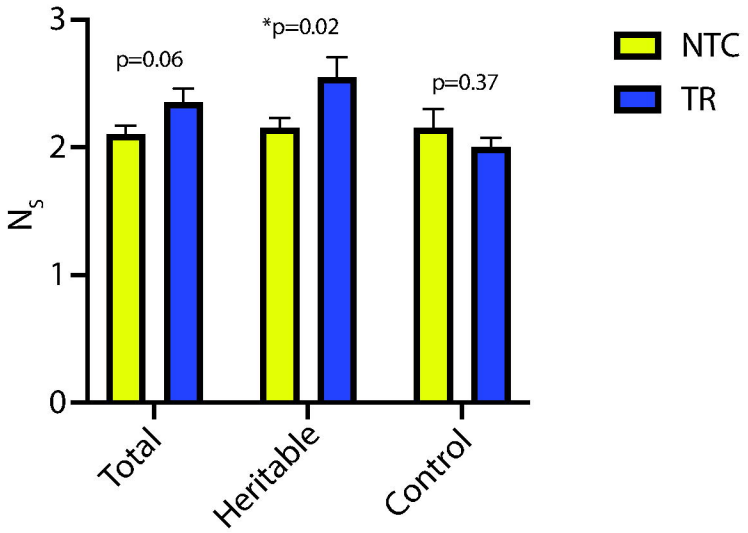
Warr, W. A. (1997). Combinatorial chemistry and molecular diversity. An overview. Journal of chemical information and computer sciences, 37(1), 134-140.


Transfer lineages (TR)
I.3
I.2
I.1
Analyze
Autoclave, incubate
Transfer 20%
Feed 100%
Feed 80%
Feed 100%
No Transfer Control pseudo-lineages (NTC)
Autoclave, incubate
Generation 0
Generation 1
Generation 2
Generation 3


300
200
100
0
#compounds enriched
0 hrs
24 hrs
autoclave
pyrite


NTC-TR
NTC
TR
* p<0.05
(NTC-TR comp. to NTC)
ms (Canberra)
0.0005
0.0004
0.0003
0.0002
0.0001
0
5
10
Generation
A
ms (Bray-Curtis)
0.020
0.015
0.010
0.005
0.000
0
5
10
Generation
B
enriched
depleted
% total compounds
30
20
10
0
0
5
10
Generation
C
r
1.0
0.5
0.0
-0.5
1-2
2-3
3-4
4-5
5-6
6-7
7-8
8-9
9-10
Generation pairs
D


complete compound dataset
every compound identified in every sample
* p<0.05 (TR/NTC)
TR
NTC
DHI, %
Generation pairs
1-2 2-3 3-4 4-5 5-6 6-7 7-8 8-9 9-10
A
"heritable core"
only compounds with r>0.5 in:
- >3/9 pairs of generations of TRs
- <1/9 pairs of generation of NTCs
B
control for artifacts
only compounds with r>0.5 in:
- <1/9 pairs of generations of TRs
- >3/9 pairs of generation of NTCs
C


3
2
1
0
N_s
p=0.06
*p=0.02
p=0.37
NTC
TR
Total
Heritable
Control