**Introduction**

Inherent in understanding the tempo and mode of evolution is understanding the variation being acted upon and the constraints limiting change. The variation of interest is often the genetic variation, or the variation in traits among parent-offspring relationships, as this is the fodder for selection to act upon. The shape and size of the G matrix is important in determining the phenotypic space a lineage can explore (site Schluter, others). If genetic constraints (i.e., covariance structures) are important, it is hypothesized that phenotypic change should occur along axes of above-average evolvability (i.e., variance). Indeed, majority of studies investigating the role of variation on phenotypic evolution show that evolution often occurs along axes of above-average evolvability (i.e., variation) (cite Voje, etc.).

Evolutionary quantitative geneticists estimate the G matrix, i.e., the amount of variance and covariance of phenotypic traits based on parent-sibling pairs, to evaluate evolutionary potential and genetic constraints on the evolution of populations and species over micro-evolutionary time scales. Quantifying the G matrix is difficult. Breeding experiments are time-intensive and often low in sample size and estimating G matrices are data-intensive (Cheverud; Porto). Further, these studies often are not conducted over deep time (cite examples including Houle 2017).

Paleontologists are restricted to estimating the P (phenotypic variance-covariance) matrix when assessing the effects of evolvability and constraints on phenotypic change within and across lineages in the fossil record. From this phenotypic matrix, a broad sense G matrix (is this right??) can be estimated. The P matrix is naturally “noisier” than the G matrix, as it represents the amount of genetic variation and environmental variation to produce the observed phenotypic variation. Cheverud 1988 proposed that a P matrix could substitute for a G matrix at sufficient effect sample sizes. This theoretical framework has been used as reason to pool phenotypic change over time and estimate a G matrix (e.g., Cheetham; Voje). There is reason to think P can substitute for G not just in the near time, but in the deep time too: the amount of phenotypic variation represented by fossil, time-averaged populations is only ~1% larger than those of modern populations (Hunt 2004?). While there have been studies showing the P matrix is a good substitute for the G matrix in some groups, this has not been evaluated using fossil data.

Clonal organisms with a fossil record, such as bryozoans, provide a unique opportunity to branch micro- to macro-evolutionary studies by disentangling the P from the G and examine changes in the G matrix over longer time scales. Bryozoans are colonial organisms with a fossil record and are a potential system to apply evolutionary quantitative genetics to deep time. Each colony is made up of genetically identical zooids. Because of their shared genetics, we can estimate the G matrix, calculate a P matrix, and extract the environmental (E) matrix. *Steginoporella magnifica* is a lineage of Bryozoa with few zooid polymorphisms and a well-preserved fossil record.

Given the potential for the G matrix to evolve rapidly under micro-evolutionary time scales, it remains unclear if the G matrix is stable enough over macro-evolutionary time scales to provide insights into multivariate evolution. Here, we apply a quantitative genetic framework to the fossil record. We first estimate G across time. We then examine how well the P matrix can substitute for the G matrix, testing Cheverud’s hypothesis. We also ask if the G matrix changes through time, and thus if the phenotypic space available changes through time. Finally, we test if, based on the directions of the G matrix, if phenotypic change is occurring in directions of above-average evolvability. Our approach attempts to apply evolutionary quantitative genetics to fossil data to investigate how the G matrix changes within a single lineage over millions of years.

**Materials & Methods**

*Geologic setting*

Fossil and modern specimens were collected from the Wanganui Basin, North Island, New Zealand during DATE. The Wanganui Basin is well studied for its geology and invertebrate fossils (cite Naish, Seward, Liow, Carter, Abbott). Bryozoans were collected from seven formations that span 2.3 MY (SI Table Formations).

*Specimen processing*

Shell substrates were cleaned using diluted bleach and water, catalogued, SEMed, and stored. Multiple, non-overlapping images of each colony were taken. The SEMS are all taken at x30 magnification, thus standardizing the scale for all the images.

*Image processing / extraction of phenotypic traits*

We applied machine learning to the SEM images to detect individual zooids and place landmarks. We piped together DeepBryo (Di Martino et al.) to detect zooids and ML-Morph (Porto & Voje) to place 23 landmarks along the zooid (see SI Fig Landmarks & Measurements) to create Steginator (<https://github.com/agporto/Steginator>). Discuss individual box id?

We only retained images of *Steginoporella maginifica* with a magnification of x30. We also deselected zooids with image distortion or errorneous landmarking (see SI for more detail?). We retained colonies with at least 5 zooids measured. Our final dataset includes X images from Y colonies (Table Sampling).

We extracted 8 phenotypic traits (see SI Fig Landmarks & Measurements). These traits were calculated as the distance between landmarks using the Pythagorean theorem (see exploratoryAnalysis.R and outputs traits.csv <- get name right). We converted pixels to um and all trait measurements were log-transformed.

Q. why outputMetadata.R?? what is in the metadata file that is useful?

*Analyses*

All post-processing of the ML pipeline and subsequent analyses were done in R programming. We used the following packages (LIST THEM). All scripts are available here (github) and all images are available here (repo).

Sensitivity

We tested for normality of each trait (see SI Fig Traits & SI Table Normality?). Include this?? We don’t seem to care…

Compared ML measurements to my own (TBD).

* Checked for quality of ML measurements compared to my own
  + 10 images from each formation
  + At least 5 colonies (i.e., all images of the same colony)
  + Placed landmarks
  + Measured self, measured using code
  + Did this three times

Rarefaction…

P & G matrix

We calculate the P matrix by first creating a variance-covariance matrix for each formation for all 8 traits. We retain the first 5 dimensions given the PC analyses (SI Fig P & G PC).

We then use the P variance-covariance matrix as the input for the G matrix prior. We then ran an MCMC glmm with traits as fixed effects and colony as a random effect to account for variation within colonies (? Rcov measurement error) to estimate G matrices for each formation. We iterated this 1.5 million times, and retained every 1000th run after .5 million runs and onward. As with the P matrix, we retained the first 5 dimensions based upon eth PC analyses (SI Fig P & G PC).

P & G correlation

We calculated random skewers following Marroig (DATE) (Figure P & G Correlation).

Change in G over time

We calculated the angle change in Gmax between time points. We get the direction of Gmax from the first eigen vector and normalize it. We get difference between Gmax as the dot product of the two matrices and covert to degrees.

? conditional evolvability…

Directions of phenotypic change

We asked if Pmax and Gmax aligned. We did this in the similar to calculating change in G over time, but rather than comparing time points compared the Gmax of time point 1 to the direction of phenotypic change seen in time point 2.

We quantified whether phenotypic change occurred in directions of above-average evolvability. To do that, we first calculate the amount of phenotypic change between each adjacent formation and normalize the vector. We then calculate the amount of observed evolvability as amount of phenotypic change based on the amount of variation-covariation in the oldest time point (t1). We then generate 10000 selection gradients in random directions to calculate the minimum, mean, and maximum evolvability for each formation, excluding the last. We compare the range of observed evolability between time points to the min, mean, and max evolvability given by the time point prior.

**Results**

P & G correlation

Change in G over time

Directions of phenotypic change

**Discussion**

**Figures**

Figure P & G Correlation

Figure G over time

Figure Gmax Pmax

Figure above-average evolvability

**Tables**

**Supplemental**

SI Table Formations. Abbr = abbreviation; Ref = reference.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Formation | Abbr. | Min  Age  (MY) | Max  Age  (MY) | Isotope  Stage  Start | Isotope  Age  End | Ref. |
| Shakespeare  Shell Bed | SHCSBSB | 0.478 | 0.424 | 12 | 12 | 1 |
| Tainui |  | 0.533 | 0.478 | 13 | 13 | 1 |
| Upper Kai-Iwi |  | 0.676 | 0.621 | 16 | 16 | 2 |
| Waipuru |  | 1.8325 | 1.826 | 67 | 67 | 3 |
| Tewkesbury |  | 1.875 | 1.8325 | 71 | 67 | 4 |
| Nukamuru  Brown Sand | NKBS | 2.017 | 1.915 | 77 | 73 | 4 |
| Nukamuru  Limestone | NKLS | 2.291 | 2.088 | 89 | 79 | 4 |

SI Table Sampling. Abbr = abbreviation; N = number.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Formation | Abbr. | N  colonies | N  Zooids  (avg. per colony) | N  Images |
| Shakespeare  Shell Bed | SHCSBSB | 50 | 400  (8) | 102 |
| Tainui |  | 19 | 155  (9) | 41 |
| Upper Kai-Iwi |  | 21 | 170  (9) | 47 |
| Waipuru |  | 15 | 156  (9) | 33 |
| Tewkesbury |  | 107 | 1050  (11) | 217 |
| Nukamuru  Brown Sand | NKBS | 263 | 2934  (12) | 525 |
| Nukamuru  Limestone | NKLS | 66 | 615  (10) | 126 |
| **Total** |  | **541** | **5480**  **(11)** | **1091** |

SI Fig Landmarks & Measurements

* Calculated the following traits based on landmarks (Figure linear):
  + Zooid height (zh) from 4 to 12
    - Similar to LZ in Voje et al. 2019
  + Median process width at the base (mpw.b) from 5 to 6
  + Cryptocyst width at midline (cw.m) from 10 to 11
  + Cryptocyst width at distal end (cw.d) from 8 to 7
  + Operculum width at midline (ow.m) from 19 to 0
    - Similar to WO in Voje et al. 2019

SI Fig Traits

SI Fig P & G PC

*Errorenous landmarking?*

SI References:

1. Carter, R.M. and Naish, T.R., 1998. A review of Wanganui Basin, New Zealand: global reference section for shallow marine, Plio–Pleistocene (2.5–0 Ma) cyclostratigraphy. Sedimentary Geology, 122(1-4), pp.37-52; Rust thesis Appendix IIIC
2. Rust thesis Appendix IIIC
3. Naish, T., Kamp, P.J., Alloway, B.V., Pillans, B., Wilson, G.S. and Westgate, J.A., 1996. Integrated tephrochronology and magnetostratigraphy for cyclothemic marine strata, Wanganui Basin: implications for the Pliocene-Pleistocene boundary in New Zealand. Quaternary International, 34, pp.29-48.
4. Abbott, S.T., Naish, T.R., Carter, R.M. and Pillans, B.J., 2005. Sequence stratigraphy of the Nukumaruan Stratotype (Pliocene‐Pleistocene, c. 2.08–1.63 Ma), Wanganui Basin, New Zealand. Journal of the Royal Society of New Zealand, 35(1-2), pp.123-150; Rust thesis Appendix IIIC