Science AMA Series: We are Hakhamanesh Mostafavi, Molly Przeworski, and Joe Pickrell, authors of a recent paper using large DNA databases to identify the ways human populations continue to evolve. AUA!

COLUMBIAEVOLUTION R/SCIENCE

Hello Reddit! We are:

Hakhamanesh Mostafavi: Graduate student in biology at Columbia University

Molly Przeworski: Professor of biology at Columbia University

Joe Pickrell: CEO at personal genomics company Gencove and professor at the New York Genome Center.

We are a few of the authors of a recent paper Identifying genetic variants that affect viability in large cohorts where we sought to use biomedical data sets to learn about mutations that affect survival.

This paper was covered in a number of news outlets with titles like Massive genetic study shows how humans are evolving, and there was a great discussion of the paper on r/science

What does it mean for humans to still be evolving? For a species to evolve simply means that mutations—the accidental changes to the genome that happen in the process of copying DNA—are increasing or decreasing in frequency in the population over time. Our basic idea was that mutations that affect the chance of survival should be present at lower frequency in older individuals. For example, if a mutation becomes harmful at the age of 60 years, people who carry it have a lower chance to survive past 60, and so the mutation should be less common among those who do. We therefore looked for mutations that change in frequency with age among around 60,000 individuals from California (as part of the GERA cohort) and around 150,000 from the UK Biobank. Across the genome, we found two variants that endanger survival in these individuals: (i) a mutation in the APOE gene, which is a well-known risk factor for Alzheimer's disease, drops in frequency beyond age 70, and (ii) a mutation in the CHRNA3 gene, associated with heavy smoking, starts to decrease in frequency at middle-age in men.We found genetic mutations linked to a number of diseases and metabolic traits to be associated with survival: individuals who are genetically predisposed to have highertotal cholesterol, LDL cholesterol, risk of heart disease, BMI, risk of asthma, or lower HDL cholesterol, tend to die younger than others. Perhaps more surprisingly, we discovered that people who carry mutations that delay puberty or the age at which they have their first child tend to live longer.

Thanks for having us, this was a lot of fun

There is a lot of talk about how natural selection is being curtailed by human medical intervention, which is the basis of the idea that human evolution isn't progressing. Is this the case?

nate

[Joe here]

Thanks for the question!

I'd say natural selection is being influenced by medical intervention, but not that it's stopping. I suspect there will be a lot of gene-environment interactions, where the importance of a gene for fertility or

lifespan (and thus natural selection) depends on the environment we live in (including medicine, etc). For example, the CHRNA3 variant that we see in this study influences lifespan presumably because people who have and smoke tend to smoke more. But in an environment where no one smokes, this variant doesn't influence lifespan.

I presume every time we change our environment (through medicine, etc.), then different genetic variants will end up being more or less important, and this will influence the path of natural selection but not the fact that it is occurring.

Hi and thank you for doing this AMA.

So if I understand your data correctly, you found that older persons are less likely to be carriers of certain disease linked alleles - in this case an AD allele and a smoking allele. Further, if you lump alleles into pathways for analysis, you reach these other findings related to heart disease, asthma and puberty.

This all seems reasonable and intuitive - older people are more likely to have fewer pathogenic variants. If they didn't they would have a hard time reaching old age.

I'm struggling, though, to understand how you go from this finding, to make the claim that this provides specific insight into how the human genome is evolving. Antagonistic pleiotropy suggests that it isn't uncommon for some alleles to be deleterious later in life, but beneficial early in life (when it matters most, from a reproductive standpoint). The classic example is something like a growth hormone - an allele which provides a more potent GH would be advantageous early in life (stronger and bigger earlier), but potentially harmful later in life (increased risk of heart disease and cancer). The puberty-delaying alleles you describe seem like they would fit nicely into this category. So why do you think these alleles are linked to the continuing evolution of the human genome - could it be that these longer lived individuals also were less Fit early in life, relative to others in the cohort? If so, then there wouldn't be any selection effects in later generations.

SirT6

Thanks for the question. We looked across the genome for variants that affect survival to a given age (for the age ranges present in the samples that we analyzed). We found only two variants with large effects (APOE4 and mutations in CHRNA3 gene), with effects manifesting late in life. If late-acting variants are not under selection, as typically assumed, or if some are actually selected for because they are beneficial earlier in life, as suggested by antagonistic pleiotropy, then many of such variants should be common in the population (unless almost no such mutations arise in the first place). That we did not observe more than two, despite having high statistical power to do so, suggests that these other variants have been kept at low frequency by natural selection (again, if they exist).

We agree that in order to understand how an allele is evolving one needs to know how it affects survival at all ages, as well as other components of evolutionary fitness, fertility and inclusive fitness. We show an example of a potential trade-off between components in the paper: variants that delay age at first birth are associated with longer lifespan, and also fewer number of children.

What are your thoughts on the recently proposed "omnigenic" model of complex disease? How (if at all) do you think this idea will change the field?

divvyflax

[Joe here]

the WINNOWER

SCIENCE AMA SERIES: WE ARE HAKHAMANESH MOSTAFAVI, MOLLY PRZEWORSKI, AND JOE PICKRELL, AUTHORS OF A RECENT PAPER USING LARGE DNA DATABASES TO IDENTIFY THE WAYS HUMAN POPULATIONS CONTINUE TO EVOLVE. AUA! : **REDDIT**

I think the omnigenic model of disease is a really provocative hypothesis, one that puts together a bunch of confusing observations from genome-wide association studies into a plausible model. Specifically I think it's a useful model to have in mind that genetic variants that influence 'peripheral'/non-'core' genes could account for large amounts of heritability simply because there are so many of them.

In terms of impact on the field, I suspect that people will think a bit differently about how to interpret (in a molecular sense) the genetic associations that they find--there's been an implicit assumption that if we push to understand the function of a gene we'll get to a clear understanding of how it influences disease risk, but that may not be the case for many GWAS hits if the omnigenic model is correct.

Hi guys, thanks for doing this AMA!

This topic is a bit out of my field, but I'm curious as to whether your research considered epigenetic factors that might play a role in longevity. I know the U.K. Biobank has a lot of datapoints so I'm just curious if this was included. Thanks!

[Austion66](#)

Thanks for the question. We didn't look at that, but could, potentially, with a similar approach.

Thank you for doing this AMA!

This is a little off-topic, but something I've thought about casually for a long time. Those of us who study "non-model organisms" tend to think of them as reflect a more "naturalistic" approach to biology. Lab mice and lab fruit flies have now lived many generations in a very artificial environment: controlled temperature, ample food, low predation risk. However, in that sense, are traditional model organisms perhaps a better model, at least for human societies in developed nations, because their environment is more controlled? Can we learn anything about present-day human evolution by examining ongoing evolution in these animal populations?

[neurobeegirl](#)

Definitely, we have learned a ton about how evolution works from artificial selection experiments on a variety of species. There are some big differences with studies of natural selection in the wild though, which may be important. Notably, if the effect of a mutation commonly differs among environments-- and there is plenty of evidence, notably from plants, that it might--then studying selection in only one environment or a limited number of environments may not give the full picture. In particular, it may be that some mutations persist in the population because while they are harmful in one environment, they are beneficial in another--whereas in a lab setting, they may be entirely harmful or entirely beneficial.

Greetings and thank you for doing this AMA.

I understand this study has focussed on adaptive selection in terms of overall health and fitness. However, I wonder if the database may also be used to look into the effects of sexual selection, which is another evolutionary selective driver. We all hear that we are living in an age of unprecedented cultural convergence where conformity to a progressively more uniform cultural preference in esthetics is dominant. I wonder if this might show in your results and how one would go about testing this hypothesis?

So: is there a way the data from your study can be used to test whether genes associated with physical appearance and esthetics are being selected differently through time (and whether globalisation is

having a measurable effect on this selection process)?

[Gargatua13013](#)

[Joe here]

Thanks, interesting question. These types of datasets can indeed be used to look at whether people with specific physical traits live longer or have more offspring (potentially due to sexual selection). In this study we didn't see any effect of variants associated with physical traits on lifespan, but one of our co-authors saw in a separate study that genetic variants that cause red hair [are associated with later age at first sexual intercourse](#).

To see whether these types of patterns change over time would be an interesting study! I think this will require a lot more data over a longer time period.

The data has been collected from US and UK citizens, so - first world countries. Are there any comparable studies done in Eurasia region, or even data available from other poor (but large, as far as population density goes) regions of the world to perform a comparative analysis in the future, perhaps?

As a plebeian with no background in this field, nor in data analysis, I feel that first-world level of healthcare and high living standards would directly interfere with obtaining (estimated, at least) results which would be relevant at a global level (the humanity as a whole). Which wasn't the scope of this study, I understand.

[Dragoniel](#)

Thanks for the comment. We agree: it is totally unclear to what extent these results will be portable across populations. As just one example, the variant in CHRNA3 will presumably not show an effect on survival in groups that don't smoke. That is actually a question we are really interested in: how often do mutations with harmful effects in one environment have no effects or even beneficial effects in another? As a first step in that direction, we can already ask how these results are affected by socio-economic status within the UK; see, for instance, http://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006711

Hi thanks for doing this AMA.

What triggers a mutation in the DNA? And how long does it take for the impact of a mutation to appear or to cause an effect on a population? Thanks

[mistymountainz](#)

Thanks for the question. Mutations occur by accident, because of damage that goes unrepaired when DNA is copied, or because of errors in the process of copying DNA. Then that mutation can spread in the population if its carrier leaves descendants, but more often than not, it is lost. The rare advantageous mutations are more likely to spread, because their carriers are more likely to leave descendants.

Thanks for doing this AMA. I assume that there are other groups involved in analyzing this large datasets---which I am sure cost a lot of money to generate. How do all these groups coordinate their research efforts to avoid issues with multi-comparisons and bias. For instance, suppose there 10 research groups, and each group studies a single independent variable. Each group computes the effect of its favored independent variable, but most variables have little to no effect, and only one

group can publish its results. Because this lone group did not consider other variables of interest they probably won't correct for the multiple comparisons being made by the whole research community. Do you think this reasoning should imply that the the statistical utility of these large datasets diminishes as additional papers are published and hypothesis are considered, even if these actions are all being taken by separate people?

felinecatastrophe

[Molly here] Interesting point. Presumably, if ten different groups asked the same question of the same data, they would get the same result. If they asked different questions, they would not be performing quite the same test. A bigger concern might be that negative results are unlikely to get published, leading to a publication bias. But hopefully there will be many fantastic resources such as these within the next few years, and then we can just see which results hold up--or if they don't, why not.

Do you anticipate you'd find different focuses for evolution in other parts of the globe and what might you hypothesize you'd find?

Do you hope to expand your data collection to other areas?

firedrops

It would be really interesting to ask these questions of other populations--and for different environments--as a big open question in the field is the extent to whether mutations will have similar effects in these different settings. As one example, the CHRNA3 variant has a discernible effect on survival in males but not females, and we think that is because males smoked more in late 20th century UK. So we would not expect that variant to show similar effects across the globe. In our analyses, we relied on big data collection efforts by Kaiser and especially by the UK Biobank. Those are phenomenal resources, but they take a huge effort to put together, and thus far, only a few of this scale have been available to researchers. We expect many more in years to come though.

Thanks for coming to talk with us! Has anyone every looked at whether socio-economic status affects these patterns? I imagine that a gene variant might have a stronger negative impact if someone has less access to healthcare?

asbruckman

This is a great point. Crude measures of socioeconomic status (SES), such as Townsend deprivation index, are usually considered as confounders to correct for in genetic studies. Indeed, as you pointed out, some genetic variants could have different effects dependent on SES, as a form of gene by environment interaction. For example, it has been reported that SES moderates the genetic influence on body mass index (https://www.ncbi.nlm.nih.gov/m/pubmed/26348497/).

How could discovery of these and other similar genetic variants translate into medical treatments? Is this a 20+ year timeline, or are we discovering actionable things?

StupidName2010

Thanks for the question. What these analyses can do already is uncover mutations that were not previously known to affect development or aging; pinpoint at what age harmful variants are starting to manifest harmful effects; and whether that differs between sexes. Also whether there are trade-offs between effects at different ages say, or potentially, across environments.