PROFESSOR: I.e. When multiple lineages of beneficial mutants are kind of spreading in a population at the same time, so when these clones are somehow interfering with each other in the population, how is it that that clonal interference effect kind of plays out in the context of these microbial asexual populations. And in particular, in this paper, what they were trying to measure was the distribution of effects of beneficial mutations. So you've put a population into some new environment. There may be some beneficial mutations that are possible. The question is, what is the probability distribution of effects of those beneficial mutations.

And this is, as you might imagine, a very basic quantitative we'd like to understand just to understand how evolution works in new environments. But it's actually something that's rather difficult to measure. And what's interesting here is it's difficult to measure for maybe a surprising reason. This is a wonderful example of how you kind of-- if life gives you lemons, then make lemonade. Is that the expression? OK. All right. So in this paper, they set out to try and measure the [? distribution ?] effects of beneficial mutations. They were unsuccessful. But they were unsuccessful for an interesting reason, so then they were able to write what I think is a very nice science paper.

Are there any questions about where we are or administrative things before we get going? All right, so just to recap where we were from last time, we introduced this Moran model, which there's a fixed population size, N. Then each of these sort of cycles, what we do is we choose one individual to divide and we choose one individual to be replaced. And we have the daughter, say, cell replace the cell that was chosen for replacement.

Now in that model, what we found was that there's some probability, x1, for a new

mutant with relative fitness, r, to actually fix in the population. And this is in the case of no clonal interference. So CI will in general be clonal interference today. So this is-- if it's just this one mutation that we're considering, then the question of surviving stochastic extinction kind of at the beginning is equivalent to the question of being able to fix in the population certainly for a beneficial mutation, with r greater than 1.

So what we want to do today to just start is to try to think about how we can use this model and this equation in order to understand the probability of fixation when there's more than one mutant in the population. So here, what we found is that this thing, there are 2 limits that we might want to keep track of. So this thing goes to 1 over n, for neutral mutations, and it goes to S for moderately beneficial mutations. So S is defined as r minus 1. So it's this for S kind of greater than zero, and for S times n, much larger than 1. So I guess this already says that S is greater than zero, so we could ignore that. Are there any questions about this before we get going to practice using it?

All right, so I just wanted to have us consider a few possible situations. All right, population size 10. What we want to consider a situation where there are two mutants. And we're going to call them A and B, with relative fitness rA 1.01, and rB 0.99. So A is a beneficial mutant, and B is a deleterious mutant. And in all these problems, we're going to assume that the rest of the population has fitness 1. So these fitnesses of A and B, they are as compared to the rest of the population. So in this case, there are eight other individuals with relative fitness 1.

So the question is, what is the probability that a fixes. Takes over the population. Reaches number of A individuals equal to 10. At that stage, we're not allowing for any new mutations here, so then that's the end of the dynamics. So you can start thinking about it, and I will give some options here.

All right so this is in the Moran model, what is the probability that A fixes. I'll give you 20 seconds to think about it. Do you need more time? Let's go ahead and vote, and see where we are. Ready. Three, two, one--

All right, we got some-- I'd say Bs, Cs is kind of dominant. All right. So I'd say that

there's generally a disagreement between B and C. Interesting. OK. I think that there's-- it's almost 50/50. It is a bit splotchy, somehow. You may need to make some effort to find somebody that disagrees with you, but please do so. All right, turn to a neighbor and try to convince them that you're right.

[SIDE CONVERSATIONS]

PROFESSOR: All right, let's go ahead and reconvene and see if anybody [INAUDIBLE]. Ready. Three, two, one. All right. I'd say that there's some migration towards C, but it's not all the way. All right, different arguments. And you may want to-- D is the argument to be made over here. All right, so let's hear it.

AUDIENCE: [INAUDIBLE]. So if there's one individual from A and one individual from B, and basically they're about the same fitness, then the probability that they should, I guess, take over is 50%. Well that assumes that one of them will take over.

PROFESSOR: Yeah, is there any reason to assume that one of them should take over?

AUDIENCE: No.

PROFESSOR: I think that we may agree that the probability that A fixes is going to be approximately equal to the probability that B fixes. That's going to be a useful insight. But [INAUDIBLE] both those things. So we have to keep track of that. So that's the argument for why it's not D.

AUDIENCE: You're welcome.

PROFESSOR: Anytime. [INAUDIBLE]. So between B and C, how can we think about this? You can tell us once again what your neighbor said. That's fine.

AUDIENCE: Why?

PROFESSOR: Yes.

AUDIENCE: Um, well, my neighbor, as opposed to me, made the correct decision to try to think about it intuitively instead of trying to get through all the math.

PROFESSOR: Yes. Always good to think about things intuitively rather than-- that's--

AUDIENCE: [INAUDIBLE]. So A is very slightly beneficial, and so it is just a little bit more likely to fix than a neutral mutation.

PROFESSOR: OK, and how can we quantify this statement? That I know I said that I didn't like math, but you know, every now and then we should look at it. Yeah, so this statement that A is approximately a neutral mutation. How is it that we can-- how do we know that? I mean did--

AUDIENCE: We have criteria for nearly neutral.

PROFESSOR: Yes, we have a criteria for nearly neutral. What does that-- what does it mean to be nearly neutral?

AUDIENCE: The magnitude of S times N is much, much less than one.

PROFESSOR: That's right. S times N is much less than one, which means that it's nearly neutral. What that means is that the probability of fixing is the same as if it were a neutral mutation, or approximately the same. And is that the case here?

AUDIENCE: Yes.

PROFESSOR: Yes. S is 0.01, so 1/100. N is 10. So S times N is equal to 0.1. Which means that both-- so A is maybe nearly neutral. Is B nearly neutral?

AUDIENCE: Yes.

PROFESSOR: Yes, so both A and B are nearly neutral. One of them is an advantageous mutant, one is a deleterious. But in this population size, they both behave as if they're nearly neutral. But in a larger population, so if we had a million individuals, would these be neutral mutations?

AUDIENCE: No.

PROFESSOR: No. So the question of whether something behaves as a neutral mutation depends

upon the population size via this relation. So that's-- and then how do we go from there?

AUDIENCE: Well the probability of a neutral mutation [INAUDIBLE] population is going to be [? 1/S ?].

PROFESSOR: That's right. A neutral mutation will fix a probability of one group. So indeed, we're going to get this. And we need to highlight that we can-- you have to be very careful about just sticking things into formulas. So you may remember that the probability of a beneficial mutant fixing in the population is S. But that's only for S times N being much larger than one, and that's when there's no clonal interference. When you don't have to compete with other beneficial mutants in the lineage. We're going to see an example of this in a moment. But if you just have a single mutant in the population, that's beneficial in the sense of it's not neutral. Then probably the fixation is S.

And of course, this is nonsensical, right? Because you can't-- it wouldn't make any-- certainly it wouldn't-- given that you say oh well, if it were neutral, it would fix with a probability of 10%. And then you say oh, but if it's an advantageous mutant, it can't fix with a probability less than that. So this is nearly neutral, fixation probability is 10%. Are there any questions about that argument?

So let's consider a different problem. So now, we're going to consider a population that's larger. So it's going to be the same situation, where we have these two mutants present as single individuals. So what we want to know now is the probability that B fixes. Once again, you should start thinking about this while I type it. Oh, there was one other thing I wanted to say, actually, over here.

If we modeled these population dynamics using differential equations-- so instead of doing the Moran process where we stochastic dynamics of division, replacement, and so forth, if you write down a differential equation describing this situation, what is the probability that A fixes? I'm going to give you five seconds.

AUDIENCE: One.

PROFESSOR: All right, one. OK, so you don't get to vote. Indeed, 1. So if you had written down the differential equation, then there's no randomness there. Whichever mutant has the largest fitness will fix with probability 1. OK? OK, probability that B fixes.

So this-- you show me both [INAUDIBLE]. I'll give you 20, 30 seconds to think about this. Yes?

AUDIENCE: Is there just one [INAUDIBLE]?

PROFESSOR: Yes. All right, so this kind of-- that's why-- yeah. So it's just the same thing. And so that's a single A mutant and a single B mutant. Single A, single B. But you should know how to calculate these things if I tell you that they were some other number.

Do you need more time? So in general, if you don't look at me when I ask that question, I'm going to take that as meaning that you're still thinking about it. Is that how I should interpret that? OK.

Let's go ahead and vote just so that we can see where we are. OK, ready. Three, two, one. All right. We're all over the place. That's great. Because that means your neighbor will either be able to help you or confuse you, or do something. So let's [INAUDIBLE]. There's a million-- population size 1,000,000. We have two mutants in the population. One has relative fitness 2, one has relative fitness of-- an advantage of 1%.

One thing to be careful of here is the equation that we have for probability of fixation here-- ah. There's one more thing that I should say. It's for S. Well, it's an S much less than 1. Just for-- so this is for a moderately-beneficial mutation. Otherwise, you have to use the full approach. Yes?

AUDIENCE: [INAUDIBLE].

PROFESSOR: All right. Go ahead and turn to your neighbor, and try to convince them that you've figured this out.

[SIDE CONVERSATIONS]

PROFESSOR: So everybody should be talking to somebody. What do you think? Well it's OK to not know, but that's why you-- you should either join them, or if you look behind you, there's someone else that looks very, uh-- All right, why don't-- let's-- OK. Let's-- what's that?

AUDIENCE: I said my reasons are always good reasons.

PROFESSOR: They're always good reasons. They'll be educational. All right, let's go ahead and reconvene. I know that there were a lot of different guesses on this problem, which means maybe that the conversations were scattered or difficult, or it was difficult to converge. But still, let's see if it made any difference. Ready. Three, two, one. OK.

All right, yeah. So I'd say that the conversations have helped. But we're still a fair distribution between maybe A, C, and D still. But-- yeah. Does somebody want to give an explanation for one of those three, or something else? Yeah, [? John. ?]

AUDIENCE: [INAUDIBLE], then you know [INAUDIBLE].

PROFESSOR: Right.

AUDIENCE: And it indicates that [INAUDIBLE].

PROFESSOR: Perfect. OK. Now I just want to make sure that we all agree on this logic. So if the only thing we had was the mutation B, then the probability of fixation would be 1% because, in that case-- is this a nearly neutral mutation? No. So that means that its probability of fixation is going to be around S, in this case, 1.01. OK, but--

AUDIENCE: But now A is there, so you-- first, you expect that-- you expect the probability of B to fix to be lower than 1%. A fixes in half of the cases.

PROFESSOR: That's right.

AUDIENCE: So in these cases, [? B's ?] not going to fix.

PROFESSOR: Exactly.

AUDIENCE: So my answer is C, but I mean, it's a bit of a--

PROFESSOR: Yeah, no, no, I mean it's-- no, but that's precisely what's happening. So for B to fix, it requires two things to happen. First, B has to survive stochastic extinction. And-- so we're going to multiply these probabilities-- A doesn't survive. We're going to discuss in a moment what this really means in terms of surviving stochastic extinction. But when a new mutation appears in the population, it will either relatively quickly go extinct, or it will relatively quickly become established, is what we call that. What that means is that it will-- if it's a beneficial mutation, it will take over unless it's outcompeted by someone else.

So it has to be that B survives stochastic extinction, and A doesn't survive this initial phase. And this thing-- so that means that the probability that B fixes is going to be-- well, OK. The probability that B survives stochastic extinction is going to be 1%. The probability that A survives is 1/2, so the probability that it doesn't survive is also 1/2.

Now the half, just to remind ourselves, comes from this thing x1, for A. There was this [INAUDIBLE]. This thing is what?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Right. So this thing very quickly goes to zero. So this is 1/2 to the millionth power. So that's zero. So we just get 1 minus 1/2. Yes.

AUDIENCE: But what are we assuming? It's not clear to me in this case. Because we're assuming they're completely independent and we can multiply--

PROFESSOR: That's right.

AUDIENCE: Together, so--

PROFESSOR: Yeah, that's right. And I would say that for-- from the same [INAUDIBLE] the population of a million, these are totally non-interacting processes, for all intents and purposes. If you were down to a population size of 10, then you might have to worry that they're going to interact more. But in this process of-- in this Moran process, you have a million individuals. It's true that this is twice as likely to be chosen to

8

divide as a normal cell, but you're talking about a million cells in that population.

So the correction to the things that you're worrying about is a [INAUDIBLE] 10 to the 6.

AUDIENCE: And so the thing [INAUDIBLE], yeah. But if rA and rB were closer, I should [INAUDIBLE] correction--

PROFESSOR: No. It's not about how close rA and rB are to each other. Even if they had the exact same relative fitness. Even if they were both 2 or both 1.01, then they would still be non-interacting. And that's just because they're in this sea of individuals.

AUDIENCE: But the thing that I'm worried about is, for example, rA could take over quickly, and then rB could take over. And that--

PROFESSOR: OK--

AUDIENCE: Would not be--

PROFESSOR: Right.

AUDIENCE: [INAUDIBLE].

PROFESSOR: OK so--

AUDIENCE: And that's not reasonable. [? If this case ?] is rA [INAUDIBLE], then rB-- the effective rB--

PROFESSOR: Right. OK So there is one problem-- You're right. So I should be careful. If the two mutants have exactly the same fitness, then if they both survive stochastic extinction, they're going to-- then you can describe everything as a differential. And this is the distinction between this initial phase of trying to become established as compared to the later dynamics in that the established means that if it's a beneficial mutation, it will spread, and basically deterministically. So in that situation, if you had two mutants, and both of relative fitness 1.01, then they both would try to survive stochastic extinction.

And let's say that they both did. Then they would both spread in the population exponentially. Of course, given that the initial dynamics are stochastic, then it would still be the case that one of them would just happen to be above the other one. But then they would grow exponentially together. That's a more complicated situation than what we're going to discuss. And that would actually be very rare. Because each one of these mutants would only have a 1% chance of surviving. So only one in 10 to the 4 times would you end up with both of these mutants surviving stochastic extinction.

Any other questions about-- so I just want to be clear here. So this is when A is equal to 1-- the number of A individuals is 1, the number of B individuals equal to 1. There was a question, what if we started with A equal to 10. So let's say the-- yeah. let's? Say we start with 10A and 10B, for example. So instead of starting with just mutant A and one mutant B, now we have 10 of each. So we're changing things symmetrically. What's going to be the probability that B fixes, approximately?

This is going to be a little funny somehow. But OK, we'll-- just think about it for 15 seconds. What's going to the probability that B fixes in this case?

All right, I'm worried you guys are now even trying.

AUDIENCE: Could you repeat the question?

PROFESSOR: Sure, yeah. OK. So we just did the problem, we had a single mutant A and a single mutant B. And we found the probability that B fixes was 0.005. Now I want to know is, does anything change instead of having a single A and single B, now what we have is 10 A, 10 B, and then the rest of them just being [INAUDIBLE] one.

So it's not fancy math. But it's still, I can see, a little confused.

I just wanted you to see where we are. We'll use the same options here, whatever is closest. And we can argue about what close means in a moment. All right, ready, three, two, one. So it's kind of some A's and B's, some E's. All right. Well, we'll talk about it.

All right. Maybe we'll jump ahead to the group discussion, just because I don't want to spend too much time on this problem. But these are the kinds of problems, I think, it's very important to practice with, because they just develop your intuition for the stochastic dynamics that occur in populations.

It's going to be something like B. You know, that's not why I put B in there. But let's just see how it goes.

And the way to find the answer here is the same as what we did before, except that now we have to think about these processes somewhat differently. So first of all, in order for B to fix, it requires that still B has to survive stochastic extinction. So we still have the same 0.01 here.

But now we have to think about the probability that A doesn't survive. And that's going to change now, right? Oh, I should be-- OK. Yeah, OK. I forgot that we had 10 here.

We could calculate what it is, and actually, maybe we should have. Well, let's calculate that. All right, so this is x1 for B. We still have 1.01 to the mill-- so we can still ignore this bottom part. But what it's going to be is, oh, no, but there's this x10 of B, where you have 10 individuals. So this is 1 minus 1.01 to the 10th power.

AUDIENCE: But we have clonal interference here.

PROFESSOR: Right now what we're trying to understand is if only we had B-- this is the probability that we're trying to calculate this part of it. So if we ignore A, what's the probably that B survives stochastic extinction? Oh, this is fine. So it's a modest change.

Because this is going to be 1 minus. So this thing is around 1.1, because 1 plus x to the n is around 1 plus nx for nx much less than 1. So this is 1.1.

OK, so right. So this is around 0.1. So indeed, I'd forgotten that there was 10. So given that we have 10 of these B individuals starting, they may survive. Ignoring A right now, just on their own dynamics, they may survive, or they may drift down to 0 and go extinct.

And indeed, even starting with 10 individuals-- 10 B individuals-- you actually still expect them to go extinct. And this, actually, calculation is very relevant for what we're about to think about, which is this question of what does it mean to be established in a population. And already you can kind of see what it's going to mean.

How big do you need to get in the population before you're likely to survive?

AUDIENCE: 1/s.

PROFESSOR: 1/s. Because you can basically see how this is going to work out in this calculation, that the number you have to get to to be established is around 1/s. So in this case, you have to get to around 100 individuals before you are going to be more likely than not to survive.

Let's see if we can figure out this problem. All right, so we had 0.1 was the probability that B would survive stochastic extinction. But then in order for B to fix, it has to not only survive stochastic extinction, but also A has to not survive.

And we can figure out what that is going to be here. This is that term. And then we need another one. So this is x of 10 of A. And this is, again, going to be 1 minus 1/2 to the 10th power.

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yeah, right. So it's always good to memorize a few of these things. 2 to the 10 is approximately 1,000. But we want not this. So we're going to subtract this 1 away. So we're going to get 10 to the minus 3.

Do you guys see how that-- right? So given that we start with 10 individuals of A with a relative fitness 2, then there's only a 1 in 1,000 chance that those A individuals are going to go extinct. So the probability that B is going to fix in this situation is rather close to 10 to the minus 4. So it's somewhere in between there.

Are there any questions about how that came about? I feel like there's a fair amount

of unhappiness.

So it could be the logic of this, or it could be the calculation of one of the two terms or it could be something else. Yes?

AUDIENCE: Sorry. Where did you get x10 B, the probably of x of B surviving again?

PROFESSOR: You want to know this term or that term?

AUDIENCE: The first term.

PROFESSOR: Right, the first term. So the first term is just asking, if there were no A, what would be the probability that B would survive stochastic extinction? That's the same as the probability that it would fix, because it's a beneficial mutation, and if B is the only thing, right? So we use this equation.

AUDIENCE: Oh, OK.

PROFESSOR: So we use this equation, except that this thing just goes to 1, so we're just left with this. But this equation is assuming that there's just one of the mutant individuals. Otherwise if you say x sub i, then it's r to the i here.

So that was the equation that was derived in chapter six of *Evolutionary Dynamics.* But I think that that's not the only thing that people are unhappy about, so please-- no? Yes.

AUDIENCE: A question on why don't you need to take into account B doesn't survive-- that probability that B doesn't survive? For that first question that's not on the board right now, I guess, for the nearly neutral [INAUDIBLE].

PROFESSOR: OK, so this was the very first one that we did? OK. Right, so you want to know--

AUDIENCE: Why don't you need to take into account the probably that B doesn't survive?

PROFESSOR: All right, so in this problem, it's just that they all behave as if they were just neutral. So the probability of each of these 10 individuals fixing in the population is the same, essentially. So they each have a 1 in 10 probability of surviving.

13

And in this case, I would say that if things are nearly neutral, you don't talk about this idea of surviving stochastic extinction, because you can't get big enough to the point where you are really guaranteed to survive stochastic extinction. Because even if you've gone to occupy 75% of the population, what's your probability of fixing at that stage? If you're nearly neutral and you get up to 75%, you have a 75% chance of fixing. In the case in the neutral mutations, there's no analogous idea of what it means to be established, because you could always just go back and go extinct with finite probability.

This question? Well, it's looking more and more likely to show up on the midterm in two weeks. So now is your opportunity to ask a question about it.

AUDIENCE: Can you talk about and establish how you got [INAUDIBLE]?

PROFESSOR: Yes. OK, this is a good question. So when we talk about a mutation becoming established, that's asking-- of course, we know that even a rather beneficial mutation, say something that confers advantage of a few percent, will probably go extinct. However, if it gets up to be a large enough fraction or a number of the population, then the dynamics are going to be well described by a differential equation where it won't go extinct.

So this is established mutation. And this, again, we really talk about only for beneficial mutations, because other mutations can't really get established. They could still stochastically fix. But it's still stochastic up to the very end.

And so the question is, at what number in the population do you had to get before you are expected to take over the population, assuming that other new mutants don't arise. That corresponds to asking, OK, this is the probability of fixation when you have i individuals in the population. And this is what we saw before.

This is permutation with relative fitness r. You have i individuals in the population, total population size n. And this becoming established means that xi is approximately equal to 1. Now, in these situations, r to the n is, again, very, very large. So this thing we can ignore. So this is really the same thing as asking that 1

minus.

And this never gets to be exactly equal to 1. But what now we're really asking is that this thing is close to 1. And this equivalent is saying that-- what do we want to say-- that 1 over r to the i is much less than 1.

And well, that means that r to the i is much greater than 1. So this is really saying that what you want is i times s to be much greater than 1. So n established, which is, in this case, it's the i such that this is true. It has to be much greater than 1 over s.

And indeed, I think that when i is equal to 1 over s, then I think the probability of extinction is 1 over e, if you do the math. So that means that this is kind of the crossing over point, where you're more likely than not to survive.

And just to be concrete, for many of the mutations we're talking about and that you read about in Roy's paper, s was a few percent. So this is saying that the population is established when the number of these mutants gets up to above 30 or a few times 30, so maybe 100 individuals means that it's unlikely to go extinct stochastically.

So really, it's once you get to this n established, then the population will grow exponentially in the population. And it's going to spread exponentially with an advantage s relative to the rest of the population.

So this discussion is actually very useful for the next thing we want to talk about. But are there any question about how we got this?

AUDIENCE: You said the probability of going extinct is 1 over e.

PROFESSOR: Right

AUDIENCE: So xi is approximately equal to 1.

PROFESSOR: Right. So this is the probably of going extinct?

AUDIENCE: Yeah.

PROFESSOR: So the probability of going extinct here, if you have i individuals, is 1 over this guy. So it's 1 over 1 plus s to the i. And then if i times s is 1 but s is small-- this is one of those definitions of e to the x that, I don't know, do you guys remember this? It's like, oh my goodness, e to the x is the limit. No, no, x is up there.

Boy, you guys are taxing me. All right. This is x over n to the n. And this limit is n goes to infinity. Is this?

AUDIENCE: Yes.

PROFESSOR: Do we agree? OK. This makes sense, right? So this is just the equivalent of saying that if i times s equal to 1 but s is small, then this thing is around e. If you're confused, I don't want to get into this too much, because it's not quite central to what we're talking about. But come up and talk to me after the class, and we'll derive this together.

So the idea is that when you've reached a population size of 1 over s, then you have a 1 over e probability of going extinct. So e pops up at all sorts of weird locations. This is why you hear about it.

So what I want to do now is I want to discuss this idea of clonal interference a little bit more, in particular ask, in what situation will we have clonal interference? Or equivalently, when can we ignore clonal interference? Can somebody say in words when we're going to be able to ignore the effect of clonal interference? Yes.

AUDIENCE: [INAUDIBLE] so large that there's a very limited match that goes to--

PROFESSOR: OK, I think we're going to have to be careful here. Because you're saying, oh well, if the population is large enough, then you're saying, oh, the mutations won't interact, right?

AUDIENCE: Not in the beginning.

PROFESSOR: Right. They don't interact at the beginning. I think that statement I'm very happy

with. But the problem is that if they both survive stochastic extinction and they start spreading, then they do start to interact in this large population.

AUDIENCE: Didn't you just say that once they survive stochastic extinction, they pretty much behave like--

PROFESSOR: They behave deterministically. But the thing is that if you have two sub-populations that are exponentially growing-- and there are these very nice diagrams that I guess in Roy's paper they don't show. Which is if you look at somehow the population kind of changing as a function of time, you can think about, OK, well, this is, let's say, constant population size. And it starts out all being a particular type. We'll say there's some population of A.

But now a new mutation arises and survives stochastic extinction. It starts spreading in the population. All right, this may be mutant B. But then if another mutant appears over here that's even more fit, then it can actually spread faster. And indeed, it can cause the B lineage to get out competed.

So we start out with population A. Mutant B is more fit than A. It's spreading exponentially. But some time later mutation C appears.

Now this C lineage is more fit than B, so it's able to spread exponentially. And indeed, it out competes B. So this is exactly what we mean by clonal interference. These two lineages are interfering with each other. So this is, indeed, clonal interference.

And there have been these drawings out. For every 20 years, people have been thinking about this. Like Lenski wrote some classic papers thinking about this. Recently, Michael Desai at Harvard has done some really beautiful work where he takes these evolving yeast populations. He does high resolution-- resolution in both senses, temporal as well as kind of deep sequencing in the population-- where he could actually directly see these lineages spreading and then being out-competed.

And he sees multiple mutations spreading through the population together because they were attached on some genome. It's a very nice paper. I almost had you guys

17

read it, so it's maybe *Nature* in 2012, '13. I don't know, possibly '12.

So this is the idea of clonal interference. So it's after the initial stochastic dynamics have played out. Yes.

AUDIENCE: Is the [? spectre ?] [INAUDIBLE] you call that [? clonal ?] [? interference? ?]

PROFESSOR: Yeah, so that is a result of clonal interference, because it's saying that that half arises because B, maybe, if it did survive, it may still get out-competed by A. So that's clonal interference.

AUDIENCE: Did you say the populations of mutations weren't [? correct? ?]

PROFESSOR: Right. So the idea is that they don't interact initially. Because when the mutations are present at a very low frequency, then they're really interacting with the bulk, the rest of the population, 10 to the 6 there. However, if they survive stochastic extinction and they start spreading in the population, then there's the possibility of clonal interference.

AUDIENCE: But wouldn't that change that [INAUDIBLE]?

PROFESSOR: No. I think that's why it's at the half. Because this factor of a half is really just saying that even if B survives stochastic extinction, it only has a 50-50 shot of actually being able to take over, because there's a 50% chance it's going to experience clonal interference with this A lineage that's going to out-compete it. Maybe if you're unhappy about that statement, harass me afterwards.

So from this discussion, though, how can we think about the importance of clonal interference? In particular, there are going to be two time scales. And the question of whether there's going to be significant clonal interference is related to which of these time scales is larger. So what might the time scales be? Or at least one of the time scales?

AUDIENCE: I guess one should probably be the time to be established.

PROFESSOR: Right. To be established.

AUDIENCE: To reach the [INAUDIBLE] [? establish. ?]

PROFESSOR: Yeah. So the actual time that it takes from when a mutation appears to when it becomes established-- I mean, there is a time there. But it ends up not being the relevant thing, because that's the regime where they're not actually interacting, that the mutant lineages are not interacting anyways. So there's a sense that it doesn't quite matter how long that takes, at least to first order.

AUDIENCE: How much time did it take to reach some population size [INAUDIBLE]?

PROFESSOR: Right. OK, yeah. So one of them is going to be that it's the time to go from established to fixation. So it's actually the next time scale there. And actually, the time to get established is actually rather short.

So a mutation appears, and most of the time it goes extinct. But the ones that don't go extinct actually get established rather quickly, because that's a biased sampling over the trajectories. But there is a rather significant time that it takes for the lineage to go from being established to actually fixing in the population.

So this is kind of how long it takes for a mutation to take over a population. We might call this the time to spread or so. And it turns out that that time-- what is it going to depend on?

AUDIENCE: s.

PROFESSOR: It's going to depend on s. So as s goes up, then this goes down. All right, perfect. All right, let's just go put a 1 over s there. And this is because it's going to grow exponentially in the population.

And then how else is it going to depend on?

AUDIENCE: n.

PROFESSOR: n, right. And indeed, this is going to be the log of n times s. It's really n over n established. Because the idea is that you start out at n established. You grow

exponentially with rate s until you kind of take over the population, size n.

And this s appears because it's 1 over n established. Because this in the log is really the population size divided by n established. So this is one time scale. Do you guys understand why it should be like this? Right. And what's the other time scale that's relevant here?

AUDIENCE: How fast mutations occur.

PROFESSOR: How fast mutations-- yes.

AUDIENCE: How long you have to wait from one mutation to the next mutation.

PROFESSOR: Right. So this is how long you have to wait for one mutation. So this is T, what we might call T mutation. Now it's a little bit subtle to think about what this should be. But it should say something about the time it takes for mutations to appear in the population.

So let's go ahead and just guess what this might be. T mutation should be equal to what? All right, so it could be--

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yeah. So I guess I'm trying to get you to think about what this should mean to be relevant for determining the time scales of whether clonal interference is important.

AUDIENCE: [INAUDIBLE].

PROFESSOR: Mu is the probability per generation of having a beneficial mutation of magnitude s. Mu is kind of the mutation rate.

AUDIENCE: [INAUDIBLE].

PROFESSOR: Right. But we'll just assume that there's only one kind of beneficial mutation, and every beneficial mutation has magnitude s, for simplicity. So it's mutation rate. It's a beneficial mutation rate per generation of leading to mutation s.

And you can just also say, if you have no idea what I'm talking about, it's fine. But it's worth spending a little bit of time thinking about what should be the relevant time scale to compare to this. Yes?

AUDIENCE: So Mu is per organism [INAUDIBLE]?

PROFESSOR: That's right. And it's OK if you find this confusing, because this is subtle. But if you think about it for 30 seconds, then maybe you'll be fertile ground for the discussion to follow.

All right, let's go ahead and vote and see where we are. Ready, three, two, one. All right. This is very nice. Almost everybody is saying C. Although it really should be D.

AUDIENCE: [INAUDIBLE]. So 1 over s is units of time, and 1 over mu is [INAUDIBLE].

PROFESSOR: OK, that's an interesting statement. I mean, I guess this is a relative fitness.

AUDIENCE: So s is mu.

AUDIENCE: But right there, you have s units of time, 1/s units of time.

PROFESSOR: Yeah, so the problem of this is this is all in units of generation time. Because this is really telling us about the number of generations. So what is going on? Because 1 over mu n should also be a time, right?

AUDIENCE: [INAUDIBLE]. Yeah, I'm not sure if-- yeah, I mean, since this is all in generations, I'm not sure if we even have to-- yeah, I think it's OK.

AUDIENCE: n times s has to be [INAUDIBLE].

PROFESSOR: I like that statement. But then he, I mean--

AUDIENCE: Well, then you have 1 over mu.

PROFESSOR: Yeah, I know. But then, this should have units of time. I agree with your statement, but I also sort of agree with his, the 1 over s here. You know that's the problem.

AUDIENCE:　　Everything is basically unitless.

PROFESSOR:　　Yeah, but I think that it's really because everything's in units of generation. So really, it's all unitless, except for mu. Yeah.

AUDIENCE:　　What I don't understand is why how beneficial the mutation is should play into how--

PROFESSOR:　　No, I agree. This is subtle. So let's imagine this is time. Now, 1 over mu n is telling us how much time there is between the initial appearance of these beneficial mutants. So here we get a beneficial mutation. But what do you think is going to happen to that beneficial mutation?

Yeah, that guy's dead. It comes up. How long do we have to wait for the next beneficial mutation?

AUDIENCE:　　1 over mu n.

PROFESSOR:　　1 over mu n. Is it going to be exactly that? Is going to be peaked around that? All right, verbal answer, what is going to be the probability distribution between successive appearances of this beneficial mutation? Ready, three, two, one.

AUDIENCE:　　Exponential.

PROFESSOR:　　It's exponentially distributed. So there might be another one that's coming through here. What do you think is going to happen to that guy?

AUDIENCE:　　Probably dead.

PROFESSOR:　　Oh, he's dead. OK. Another one. So these guys are appearing. But if the magnitude of s is, say, 0.02, that means we have to wait for 50 of these things to appear before we expect that one of them is actually going to get established. And established means it gets up to 1 over s. So this guy, once he's established, he's going to grow exponentially in the population.

But the idea is that the time scale here of 1 over mu n is the time between successive appearances of this beneficial mutation, whereas as 1 over mu ns is the

time scale between successive establishments of these beneficial mutations. And it's only if they get established that they're relevant.

So that's why the t mutation established is 1 over mu ns. Now, no clonal interference means that one of these is larger than the other one. So no clonal interference means is it, A, this guy's larger, or B, this guy's larger?

So larger, question mark. This one? That one? All right, think about this for 15 seconds, just to make sure that we're-- make sure. We're asking, if there's no clonal interference, it means that one of these is much larger than the other one. Which one? Ready, three, two, one. It means that this thing is much larger than this one.

So this thing has to be much larger than t establish to fix. Because if it's the case this time scale is very large compared to this time scale of spreading, then the mutations don't ever clonal interfere.

I think it's very important that you can reconstruct this whole argument because it has all the ingredients of the dynamics in terms of stochastic, and then deterministic. If you find this stuff confusing, please go through it with a friend or come to me after class.

So what we went to do for the last 20 minutes is, then, talk about Roy's paper. And I think that this paper is interesting and subtle. It's a little bit hard to tell how much of it is an experimental kind of straight up demonstration or how much of it is that it's just a way of getting you to think new thoughts.

It certainly is a way to get you to think new thoughts. At least I felt that it had that effect on me, and that after reading it, I just got very excited about all the dynamics that were at play here, and how out of the complexity of this clonal interference process, maybe, in some cases, there might be some even simpler phenomenological type description of what was going on. But I think that what is very interesting about that paper is this idea that what they really wanted to measure was this probability distribution of beneficial mutations.

So what we mean by that is that different beneficial mutations will have different

effects. And in particular, you could imagine a thought experiment where you take an E. coli cell. How big is the E. coli genome roughly?

And just like it's always good to have guideposts for how long ago things happened, it's also good to have guideposts for things like it's just good to have a few genome sizes memorized. Now does anybody know the E. coli genome?

AUDIENCE: 2,000 base pairs.

PROFESSOR: 2,000 base pairs?

AUDIENCE: 2,000 [INAUDIBLE].

PROFESSOR: No, I just want to make sure that we're-- yeah. Yeah, it's a bit more than that. It's a few million base pairs, three or four, depending on the strain.

But you can imagine a thought experiment where you go in, and at every base pair you make the three possible point mutations. So then you can imagine having 10 million different strains that are each different at one site. And then what we can do is we can plot the rate of population growth in some environment.

So this is gamma. It's the 1 over n dn/dt, basically division, right? Normalized by the wild type division. I'm sorry, that's supposed to be down there.

So this is gamma over gamma of wild type. And this could be E. coli in LB or minimal media or at 30 [? ci ?]. You pick some reasonable environment.

Question is, if you draw the histogram of what this thing should look like, what should it be? Now what I'm going to ask you to do is, in 30 seconds, draw something on your sheet of paper. This is supposed to be a frequency or histogram over those 10 million point mutations.

So we haven't actually done this experiment. But I think we have made all possible gene knockouts in E. coli, so removing each of the genes. And we've done that in E. coli. We've done it in yeast.

And indeed, in yeast, they're actually making and measuring the division rate for all pairwise knockouts. So they're some fraction of the way through. They published their first 10 million measurements for the division rate of the pairwise knockouts in yeast. This is Charlie Boone at Toronto. It's an amazing data set.

Sorry, this is a simpler one. So this is just if we make a histogram of the growth rate of E. coli, all 10 million possible point mutations in E. coli, what do you think it's going to look like? And I'm going to come by, and if I don't see distribution drawn on your sheet of paper, then you'll get to draw it up here on the board for all of us.

AUDIENCE: It's totally a distribution.

PROFESSOR: OK. All right.

AUDIENCE: He was coming. He was coming to me.

PROFESSOR: All right. And what I find interesting about this exercise is that you get all possible distributions. No, but it's funny, because it's like a super basic question. And somehow we don't necessarily think about it, or whatnot. Could somebody throw out what they think maybe should be going on? Yeah.

AUDIENCE: So I'm expecting a large fraction to get 0.

PROFESSOR: So it could be that they're all here.

AUDIENCE: You know, most will start-- not most, but large fractions [INAUDIBLE].

PROFESSOR: So it could be that a large fraction of these point mutations, and the cell is just dead. So this is what you call a pessimistic view of nature or of life. And he has trouble crossing the street because he's never sure. What do you think?

AUDIENCE: I actually think that there's a much larger fraction where it's exactly the same, because they've got a lot of point mutations that are just substitutions.

PROFESSOR: Yeah, OK. So lots of them, maybe. So these are the two polar views of the world, that nothing matters, or you're all dead.

25

AUDIENCE: I think it can reconcile it. I think you can say--

PROFESSOR: OK, well, now I think you're being too-- OK, how are we going to reconcile them?

AUDIENCE: So we know that there's the [? third ?] [? bubble, ?] right?

PROFESSOR: Sure.

AUDIENCE: And so there's probably at least 1/3 somewhere it just doesn't matter at all, where you get the same code.

PROFESSOR: Although there are cases, actually, where even silent mutations end up changing protein at an expression level. All right, but I don't know if this reconciles the two views of the world.

So some people have drawn things that look like this. Some people will draw it, and things look like this. Yeah. And we have a uniform distribution. If you're going to make a null model, I mean, I don't know.

So I would say that we have enough information to say something about what this thing should look like. So first of all, some fraction of genes are indeed essential. For 10% or 20% of the genes, you remove it, the cell is dead.

That doesn't mean that 10% to 20% of point mutations will lead to a lethal phenotype, but it means that if you do inactivate or knock out that gene, then it will be lethal. So that means that there are indeed some set of these point mutations that are going to be lethal, but it's going to be small.

And I don't know what the actual number would be, but maybe 1 in 10 to the 4, something small. Because it's probably only 10% the locations on the gene would actually knock it out. I'm making up that number. The protein guys would have a much better sense. But it's going to be a small number. It might only be 1% of the mutations, actually, would actually knock out the function of the protein.

And then not all of the genome actually codes for proteins. So you start multiplying small numbers together, and you get something that's rather small. But there will be

some number of mutations there.

However, the vast majority of mutations will have no measurable fitness effect. So indeed, this thing kind of peaks here, and then rather sharply comes down. And this is on a linear scale, or something like that. So the width might indeed be a few percent. But actually, probably even less, maybe 1%, or half, something small. Yeah.

AUDIENCE: How skewed is it?

PROFESSOR: So on a linear scale, it doesn't look skewed. On a log scale, I think it does. I mean, in the sense that if you plot log frequency on this axis, then there's a longer tail on the left than on the right. But I'd say on a linear scale, it's just pretty sharp function there.

AUDIENCE: So people have measured that?

PROFESSOR: Certainly for all the gene knockouts, we have measured this. And even that distribution is sharply peaked. So the point mutation distribution's going to be even more tightly peaked.

AUDIENCE: For most genes, you can just knock them out and [INAUDIBLE].

PROFESSOR: Yeah. Sorry. So for most genes, at least in many environments-- and of course, it depends on what environment you measure. But for most genes, you can knock them out. Certainly in rich media, most genes you can knock out.

This, as I said, might be 1 in 10 to the minus 3, 4. I might be off by a couple orders of magnitude. But the point is that this is a small fraction over here. Most are here.

And there will be some point mutations that come down here. But on a linear scale, it's going to be very small. Now from the standpoint of a population evolving to a new environment, which mutations are we most interested in?

AUDIENCE: The beneficial.

PROFESSOR: The beneficial mutations. So we're most interested in these guys. And indeed, this is the distribution that [INAUDIBLE] set out to try and measure in that experiment that we just read about last night. So he wanted to know, if you zoom in here-- because remember, this is even narrower than I've actually drawn. But this distribution, the probability of a beneficial mutation of magnitude s-- s, here 0, s-- it's going to do something. Many mutations that are nearly neutral, but it's going to fall off in some way, and they wanted to try and measure this.

Now how many mutations confer a 2% advantage, 4%, for this particular E. coli strain in that particular environment? But it turned out being difficult to measure. And what was the reason that they gave for why it was difficult to measure? Somebody? Anybody? Please?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yes. But the equivalence principle doesn't say that they're all equivalent necessarily.

AUDIENCE: Well, you can make different [INAUDIBLE]. So there are different distributions, and they each have their perimeters

PROFESSOR: Yeah.

AUDIENCE: You can pick perimeters for each of them, such that they will give you the same--

PROFESSOR: OK, that's right. So what they found was that different underlying distributions-- this probability distribution for beneficial mutations, function of s-- different underlying distributions could give you the same final output in their measurements. But why was that? Yeah, John?

AUDIENCE: The only thing you can measure is what fixes.

PROFESSOR: Right. So you can only measure what it fixes, or becomes significant fraction population, 10%. Only measure what fixes or grows. And what does that mean?

AUDIENCE: Well, it means that you're not probing [INAUDIBLE].

PROFESSOR: Right. It means that we're not probing that distribution, and why not?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Yeah. But why is it that we only see part of this distribution? What was their--

AUDIENCE: We don't see the ones that die out.

PROFESSOR: We don't see the ones that die out. And why is that they die out? Or why is it that they--

AUDIENCE: [INAUDIBLE].

PROFESSOR: OK, yeah. Yes. So the fact of stochastic extinction is very relevant here. And I think that in this paper they a little bit underplay that aspect of it, because they're primarily arguing that it was another effect that led to this equivalence principle. Right?

AUDIENCE: Competition.

PROFESSOR: Yeah, clonal interference. I mean, that's why we spent the last hour talking about clonal interference. So their argument in this paper was that, as a result of clonal interference, competition between these different lineages, then the distribution of fitness effects that you measure, that fix, is not the distribution that was the underlying distribution. Because less fit beneficial mutations get out-competed.

But there's a very important question, which is, let's imagine that let's have mu go to 0. Or we go to small populations. Let's say that after reading his paper I said, OK, well, yeah, because of clonal interference I can't measure this probability distribution in the set up that he used. But maybe if I go to small population sizes, or I reduce the mutation rate somehow-- magic-- then would this allow me to measure that distribution?

So I'm going to ask you guys to vote. So mu, let's say. Our experiment, I have mu go to 0. Does this mean that when I then go do write CFP, YFP, they're 50-50, and then you start doing this. And we measure the slope to get s, just like in this experiment. If I plot the distribution of resulting s's, no clonal interference anymore,

29

because I've set mu equal to 0. So this thing goes to infinity.

Do I recover the underlying distribution? Ready. So P be measured, as a function of s, is it equal to the true one? Question mark. I'm going to give you 15 seconds to think about this.

Ready. Three, two, one. All right. So we have, I'd say, a majority of nos. And can somebody say why not?

AUDIENCE: You can't really see the established mutation.

PROFESSOR: That's right. Because even if you don't have clonal interference, you still only see the mutations that survive so stochastic extinction. And I think this is a really important point that, at least in a first reading of the paper, you might not realize. Yeah, John.

AUDIENCE: [INAUDIBLE].

PROFESSOR: That's right. So in principle, we know that the probability of survival goes as s. But you end up losing a fair amount of information. So for example, let's say that this thing is an exponential. So this is Pb of s. Let's just say it's an exponential. And indeed, from some ideas from extreme value theory, there are arguments that maybe this should be an exponential. And if you're curious about this, come and ask me after class.

But let's just imagine that this thing is exponential. Now the question is what would we actually measure in this experiment in the absence of clonal interference? Would we measure a distribution-- would the distribution be maximal at 0 or at some other value? Verbally, 0 or other value? Ready, three, two, one.

AUDIENCE: Other value.

PROFESSOR: Other value. And indeed, the probability of surviving extinction, remember this $x1$ goes as s means that the distribution they actually measure will start at 0, grow linearly, and then fall off. So this distribution is essentially like this distribution of the convolution of two exponentials, where this thing goes up linearly, and then curves

30

off.

So I think that a very important point to realize in all of this is that this is probability distribution that's measured as a function of s with no clonal interference, right? So the statement is that even if you don't have any clonal interference, already you're going to measure a peak distribution. The underlying distribution will be peaked at 0.

But what you measure is always going to be peaked somewhere else. And indeed, if you go and you look in the paper at the measured alpha as a function of s, you see something that looks kind of like this. So it's a little bit difficult to be absolutely certain.

What's going on is that in the absence of clonal interference, you get this. With clonal interference, let's say you have two mutations and you want to know, OK, what are you going to measure? Well, you sample from this distribution twice, and you take the larger of the two. Because this distribution is already the probability distribution of mutations that get established.

So then this is going to grow as a quadratic here. If you have more clonal interference, then this thing shifts to the right and gets tighter. So this is the process by which you end up with a peaked distribution.

So what we're going to do is we're going to start in the first 15 minutes on Tuesday. We'll talk about this experiment a bit more. And then we'll go on to think about fitness landscapes and the rate of evolution. If you have any questions about anything that we said, please come on up.