Katarina Piponi: 
Hi, I am extremely excited to be part of the Biodiverse Festival 2020. And I am here to talk to you about a novel derivative of Aequorea victoria's green fluorescent protein, otherwise known for short as GFP. So I'm Katarina Piponi, I am currently in my fourth and final year of my MSci Zoology degree at the University of Nottingham, and I've immensely enjoyed my degree so far. But most importantly, it led me to last year start evolving green fluorescent proteins in the lab and discover the derivatives that I named mGFPwt. I have also really enjoyed researching, reading and discussing so many fascinating topics within my degree. It has inspired me to devote my scientific career to researching evolutionary ecology and conservation. I thought well, I should start off by introducing the protein. So here to the left, we have GFP well it was first isolated from the jellyfish Aequorea victoria in 1962. An array of fluorescent proteins followed with a rainbow of colors, molecular structures and excitation dynamics. And these all these molecular discoveries have propelled their utility in scientific research as biosensors, dyes or tags, which have all allowed us to monitor cellular processes in the living systems. The green fluorescent protein is made up of 238 amino acids folded into a beta-barrel shaped protein with a folded alpha helix in the center. And here, if you look at the crystal structure to the left, and you can see that the cyan represents the beta strands, magenta represents the helical region in the center, and the beige represents the loops. And it's within this magenta part, the helical region sits the chromophore tripeptide. This is a very unique light emitting region. And I'll tell you more about the chromophore tripeptide when I explained the impacts of my derivative's mutations, and I should point out actually that this protective beta-barrel the cyan part, results in the chromophore being extremely self sufficient, requiring no specific enzymes from Aequorea victoria and this makes GFP derivatives almost ideal engineering and transforming tools, and it was because of this that we could insert synthetic DNA coding for GFP into vectors and the fluorescent protein would still emit fluorescence.

So what is my research? Novel fluorescent proteins have been synthesized and discovered continuously since 1962. However, fluorescent imaging of transformed bacteria still use bright GFP derivatives that due to trade offs impede recombinant bacterial growth. So while evaluating GFP derivatives post mutagenesis, I identified that there was a bright derivative, but that it also grew larger colonies. In overview, GFP derivative, which I named mGFPwt was identified after applying error prone polymerase chain reaction to GFP in order to gain random mutations in the sequence. And we then recognised and selected it after viewing it on a plated recombinant E.coli in blue lines. And it was then sequenced and analysed, where particularly interesting differences in the amino acid sequence identified some mutations that were on the surface of the beta-barrel and or close to the chromophore, arguably leading to the brighter emission and less impact on growth.

So how did I do it? I should start off with explaining how did we see our fluorescent proteins in action. And what we did was we used vectors, we used vectors constructed from the PQE30 plasmid, and you can see a schematic here for it to the left, which we actually inserted into E.coli. What was key here was that by inserting synthetic DNA encoding for GFP upstream of a red fluorescent protein, otherwise known as mKate2, we could control for the fluorescence intensity per cell when comparing bacterial colonies. After applying mutagenesis, GFP derivatives were expected to have differing fluorescence intensities. And we didn't want our analysis of this to be affected by the size of the colony, which would have a brighter fluorescence the larger it is. Therefore, by including a consistent and unmutated mKate2, we were able to measure novel fluorescence as as a ratio compared to mKate2's emission with the wild type GFP, the unmutated GFP.
Okay. This then leads me to answer how did we mutate and clone our GFP derivatives that we then inserted into our PQE30 plasmids. To mutate their DNA sequences, including GFPS were amplified under error prone polymerase chain reaction conditions, which act to destabilise the reaction buffer composition, and cause Taq DNA polymerase to make low fidelity copies, or simply just to make mistakes, therefore, applying mutations. So it would make mistakes when copying the DNA sequence. And importantly, here the number of ep-PCR cycles were kept the same, so that variants in template stock lead to variants and Taq DNA polymerase readings. What that means is less template increased reading, and therefore errors in the matched nucleotides. So as a lab, we decided to use three mutation rates, a low, medium and high mutation rate and after analyzing mutagenesis results, we saw on average that that paired one and about 1, 5, and 10 mutations per template. Interestingly, which you can see here in the diagram, when we electroporated, the ep-PCR or error prone polymerase chain reaction products, we could see the impacts of the mutation rates on the DNA fragments, the more mutations applied, so the higher the mutation rate, the more fragments were cut short and discarded. So we had less exposure, we had less products, and we had more primer dimers. I found that even more interesting that with so many mutations applied, even the high mutation template products still showed a fluorescence. So I decided to splice the high mutation rates products, to see which mutations were potentially influencing the fluorescence emissions. After extracting and separating the GFP derivatives from the electrophoresis gel, we needed to transform our chemically competent E.coli, and we decided that the Gibson Assembly protocol was most effective here. This was just because it was a new faster format of molecular cloning that didn't require restriction enzymes. And you can see this in my schematic here where our three components, the Gibson Assembly Master Mix, the DNA inserts and the vectors are all deposited into microcentrifuge tube, then you can follow my diagram, where you see then a close up of the microcentrifuge tube, which shows the scarless ligation method occurring while subjected to heat shock. And finally, the final solution is plated onto Luria broth medium, which contains 0.1% ampicillin and that was just a control to make sure that the colonies that grew were the ones that actually took up the plasmids. And that was because we made sure there was a ampicillin resistance gene in the plasmids. You can see the products of our plating here, it was identified that the colony containing mGFPwt, pointed out here by the cyan arrow in the centre plate alluringly avoided detrimental impacts on E. coli growth. You can see that because the mGFPwt colony grew extremely large colonies, similarly to the vectors transformed with a six histidine tag instead of an additional fluorescent protein, that's just to the left. And it grew much larger colonies than the wild type GFP colonies on the far right. And what's interesting here is that the GFP sequence has high structural malleability and high photo physics sensitivity of the chromophore to a broad spectrum of physiochemical factors. Therefore, it comes as a surprise that the mGFPwt transformed in E. coli plasmids had low impact on bacterial growth. While still the chemical environment and the folding of the surrounding protein matrix will still able to produce an appropriate domain for fluorescence and particularly interesting a brighter florescence. So taking into account colony sizes, I decided to replay 12 of my most intensely fluo rescing colonies, and I selected these using a relative pixel intensity threshold after imaging them and the left plate here has the mGFPwt colony in the cyan box and was among 11 other colonies that I wanted to explore further which are pointed out by the white boxes. And then I replated these 12, which you can see on the right hand diagram B where again, the cyan represents the colony containing mGFPwt. The enlarged close up here on the right diagram is colorized to indicate the red to green fluorescence intensity and I created this by layering two images. One image was of the red channel and the other of the green channel, and they were both taken with the same capture settings. And so the orange tone of the
mGFPwt economy suggested mKate2 is performing highly efficiently. And that suggests a potentially less impactful GFP structure within the plasmid. But, before we sequenced our novel GFP derivatives, we had the entire UV vis spectrum of the samples captured to explore the fluorescence further, the plate reader detected chromophore emission by exciting electrons from their ground state into an excited state. So, due to always analyzing the fluorescent proteins as part of the E. coli plasmid, an optical density of the samples at a wavelength of 600 nanometers, otherwise known as OD 600 were recorded, recorded in conjunction with artificially emitted illumination. This sets a control for bacterial growth.

This was because we expected the bacterial growth to influence light emission. The OD 600 absorbance measure calculated the concentration of bacterial cells suspended in the wells, a higher concentration of bacteria will deflect more light from the detector beneath the well plate. So, after taking OD 600 measurements into account, we saw a significantly more intense green emission and a minimal non significant change in the red emission mGFPwt. The GFP wild type is shown here on the left with the mGFPwt beside it and the following one to eight GFP derivatives are from the highest performing colonies on the green channel for my lab, just for comparison, and the PPS and LB are averaged controls for Luria broth and phosphate buffered saline from well plates that contained two solutions only. And that was just to control for that interaction with the emitted light while they diluted and suspended the bacterial cells. And the error bars represent 95% confidence intervals. And what is really interesting here is a greater intensity was obtained and didn't knock out the red fluorescence or impede bacterial growth. Contrastingly, experiments done by Rang et al. 2003 and Han et al. 2015 saw brighter green fluorescence proteins have an inverse relationship where the growth speed of gastrointestinal bacteria slowed. And as a result, the fluorescence intensity was described as a fitness cost. So finally, we ran our sequencing and what I found sparked a lot of interest.

After running a multiple sequence alignment with the modified GFP derivatives, eGFP cycle three and sfGFP, which are super folding and enhanced derivatives, I found that mGFPwt had 17 amino acid differences with the wild type. Above the sequence alignment here, I've labeled the crystal structures of the protein to match the crystal structure on the right.

And I have also provided a close up of the helical region where in yellow, you can see the chromophore tripeptide and this is the light emitting region. So I've circled the sequence alignment here that you can see the glutamine to arginine mutation at position 85, which is remarkably shared with sfGFP and cycle three. The modified alternative sfGFP is a superior folding derivative superfold with GFP, that is scientifically applied for its protein stability and rapid maturation. The cycle three mutant on the other hand adjusted its folding from just three surface mutations and avoided any effect on the stability of the protein, but rather resulted in less aggregation due to a substitution with hydrophilic amino acids. So it's really fascinating to see a mutation in common. However, thus far neither sfGFP nor cycle three have been reported to avoid impacts on bacterial growth. So, the unique characteristics of mGFPwt demonstrate potential novel routes for engineering fluorescent proteins. And here are the, on the multiple alignment again, I have circled a cluster of mutations between residue numbers 17 to 30 on Beta one and Beta two strands, and this cluster causes hydrophilic amino acids lysine 20, arganine 24, thanine 28 and asparagine 29 to protrude outwards of the Beta barrels and in the same region, an mGFPwt has a prolene inserted, proline residue inserted between residues 30 and 31. And considering the lysine insertion at residue 20, these are major disruptions to the surrounding
barrel. The prolene actually sits in a conserved residue position which has shown roles in determining the final folding in the GFP derivatives. And sfGFP interestingly, was constructed using a mutation in this position arginine 30, which was found to cause the greatest improvements in folding robustness and refolding kinetics. And similarly, the prolene my protein insertion in mGFPwt at residue 30 does cause local conformational changes. It effects side chain hydrogen that side chain hydrogen networks of cyanine 30 and theanine 48 by disrupting a hydrogen bond between them and this disruption causes a destabilization which I theorize is loosening the loop above the Beta one strand.

And most intriguingly, residue 151 proline replaces histidine and therefore no longer acts as a nitrogen donor to tyrosine's oxygen acceptor at residue 71. This means that the mutation at residue 151 destabilizes the chromophore state at tyrosine 71 residue. And that was a phenomenal find. Therefore, it comes as a surprise that the dissimilar polar contacts between the chromophore tripeptide and the hydrogen network results in a brighter performing green fluorescent protein, mGFPwt and I definitely think that this is potentially responsible for mGFPwt's brighter emission. And this is the structural impacts of these mutations. You can see them in a modelled overlay of mGFPwt in cyan, and GFPwt in grey. So, the mutated one is in cyan and the wild type is in grey. The crystal structures are shown below and the surface figures are shown above, you can actually see the loosening of the loop above Beta one is now visible, and the Beta one strand shift is strikingly visible. To see that even better, I've isolated the strand and enlarged it here to the right. And it's the relative positioning of Beta one strand is certainly the most striking difference between the two. It's been understood that specific bits of our folding acts as a prerequisite for the initial, well, the initiation of chromophore maturation, and bending of the central alpha helix exposes the inner tripeptide, the chromophore tripeptide, and arranges catalytic amino acids in positions that are favorable for synthesis to perform the chromophore emission. In mGFPwt, the Beta one strand shift pushes isoleucine 16 towards valine 73 in the alpha helix, were originally a valine residue would have protruded. This is potentially an interruption to the interacting aqueous environment resulting from the addition of a less hydrophobic side chain facing the inner alpha helix. And these mutational impacts are novel design to the scientific community. And therefore, excitingly, I think they hopefully inspire new roads for research.

So what does this mean for science? The discovery of novel GFP derivatives have many implications. Crucially, that implication in science elevates fields across the board from immuno therapeutic treatments to mapping the brain's neurone activity. Ultimate, the discovery of mGFPwt boosts the experimental manipulation or monitoring of bacterial cells and colonies, particularly in the gastrointestinal regions, where bacterial growth is a valuable feature of the data, and requires the least impactful tag. Equally, the impact of mGFPwt on bound mKate, expand the potential for new roads or fusion type developments, given the potentially low impact on a variety of recombinant proteins. In addition, these phenotypic findings set up with corresponding structural analysis, deepen our understanding of the mechanisms used to trigger chromophore maturation, and maintain fluorescence emission. On the basis of our modeling, it appears the supporting hydrogen network surrounding chromophore is more adaptable than first thought, and although further studies are required to unravel the bio technological specificities of mGFPwt as a bacterial or fusion tag, these discoveries provide new opportunities to further explore the functional consequences of mutations in the GFP sequence.
And that concludes my talk. I hope you enjoyed and potentially learnt something new about fluorescent proteins and thank you for listening.