QUEENSLAND FRUITFLY

Contents

1. INTRODUCTION

Queensland fruitfly (Qfly), *Bactrocera tryoni*, is a key agricultural pest along the East coast of Australia. The current genus name Bactrocera is of quite recent standing, following the taxonomic revision by Dick Drew in 1989 [2]. Prior to this Qfly was known as *Dacus tryoni*, which causes some confusion as a substantial amount of the literature has the old name.

The School of Biological Sciences at Sydney University has a long history of involvement with Qfly. It was the principal organism of Charles Birch, the Professor of Zoology for many years. During his time the CSIRO Division of Entomology under Alan Bateman established a section within the School to work on Qfly..

Most of this work was behavioural and physiological. The only genetical work that I know of was done by Steve McKechnie in a PhD with Charles Birch, using isozymes identified by electrophoresis, mainly to study the difference between *B. tryoni* and a closely related species *B. neohumeralis*. There had, in fact, been a bit of work along genetical lines by Birch and Dick Lewontin, proposing that it was hybridisation between the two species that had led to the adaptation of *B.tryoni* to cooler climates.

1.1. **Founding of the Fruit Fly Research Centre - FFRC.**

Sometime around 1990 I was thinking of the possibility of branching out from Drosophila work. I have to admit that this was partly because of the difficulties of continued funding for Drosophila work, and I wanted to look at the possibility of working on something for which commercial funding might be available. Maybe there was also a desire to do something a bit more applied and/or useful.

I was quite struck at the time at the differences between the approaches taken for Qfly and for another Australian fly that was the subject of much work in CSIRO and universities, the sheep blow fly *Lucilia cuprina*. There was a huge amount of genetical work on the latter, contrasting with the low emphasis on genetical work in Qfly, mainly

owing to the scientific background of the people working on the two species.

Around this time Marianne Frommer, my wife, was looking at the possibility of moving out of her work in a cancer lab on the other side of Sydney and moving to Sydney University. Marianne's background was in molecular biology, exactly what was needed at the time for moving into genetical work on a new organism. She put in a proposal for an ARC Senior Post-Doctoral Fellowship to work on Qfly, which, somewhat to her surprise, she was awarded. In retrospect Marianne made one of the worst career moves of all time. Just prior to moving to Sydney University she published the bisulphite method for sequencing of methylated DNA [4], which has had an enormous impact in screening for pre-cancerous states and would have made her rich and famous had she remained in the field.

Neither Marianne or I knew anything much about Qfly. Fortunately a third person, Alfie Meats, was already involved in fruitfly work at Sydney University, and knew almost everything about Qfly. So the combination of the three of us, with our respective backgrounds in population genetics, molecular biology and entomology seemed ideal for the project.

I should also mention a fourth person, Merryl Robson, who had been a fellow undergraduate at Adelaide University. Merryl's great strength, aside from her genetics knowledge, was in fund-raising. She managed to persuade Woolworths to give us seeding funds, which together with matching ARC funds enabled the FFRC to prosper for a number of years. I always hoped that Woolworths got something from the project, mainly in terms of kudos from the growers with whom they worked.

For a history of the various activities of the FFRC see https://qffrl.wordpress.com/

2. STUDIES ON THE DISTRIBUTION OF QFLY

One of our first aims was to develop genetical markers to be able to differentiate populations. Literally nothing was known about this, except that there was a large base population in northern Australia, mainly Queensland, and smaller populations in NSW and Victoria. Historical documents showed that Qfly had arrived in Sydney around the beginning of the 20th century.

As might be expected, one Qfly looks pretty much like another. Steve McKechnie's isozymes hadn't shown anything at this level. Even physiological studies to show cold tolerance of flies raised from the south as opposed to flies raised from the north were not revealing.

There were a few possibilities for genetical markers at the time. The method on most people's lips was RAPDs, Random Amplification of Polymorphic DNA. This method relied on fingerprinting of flies chosen using fortuitous PCR primers that showed up differences between strains. It was quite easy to set up, and perhaps for this reason very popular at the time. However the variability between samples could be a problem.

Less well known at the time were 'microsatellites'. These are short regions of repeated dinucleotides or trinucleotides that differ between individuals in the number of copies of the repeat. Each such microsatellite needs to be individually amplified by PCR using specially chosen primers. Marianne had the insight to see that despite the much greater amount of work needed to set up the microsatellite system, it had huge advantages in terms of reproducibility when used on the rather degraded DNA samples that one gets from trapped flies. So we started on the right track at the beginning. Nowadays microsatellites are allpervasive in the field of population genetic studies.

Our original isolation of microsatellite markers produced six good polymorphic markers [6]. This was somewhat less than we had hoped for, but, fortunately, turned out to be enough to get very useful information.

2.1. **The CSIRO Double Helix study.**

Getting samples is always a problem in field studies of any organism. Fortunately we found out that the CSIRO Double Helix Club to involve children in science was looking for a national project in the year 1994. It took a lot of setting up, but the project turned out to be a triumph. Children were shown how to make a simple trap using a cut soft-drink bottle. A lure that attracts male Qfly and some insecticide were sent to them in the post. They made observations on the weather and the number of flies trapped over a 2-week period, and then sent us the flies preserved in a small bottle containing alcohol.

Many hundreds of samples arrived at the lab. The flies were sorted into species by Rachel Osborne, an Honours student in the lab with an entomology background [10].. The Qfly, the majority species in most cases, were deep-frozen.

The original 1994 sample involved all children in the club. Some from the north trapped hundreds of flies. Others in southern regions collected none (fortunately), although I hope that they felt some satisfaction in being part of the overall project. A subset of the children who had provided the samples from regions where Qfly is endemic were asked to remain part of the project in 1995, and the study was extended up till 1998. In some cases parents took over the sampling, and continued to send us samples beyond that point. We ended up with good samples from several regions for the years 1994 - 1998 (Figure 1).

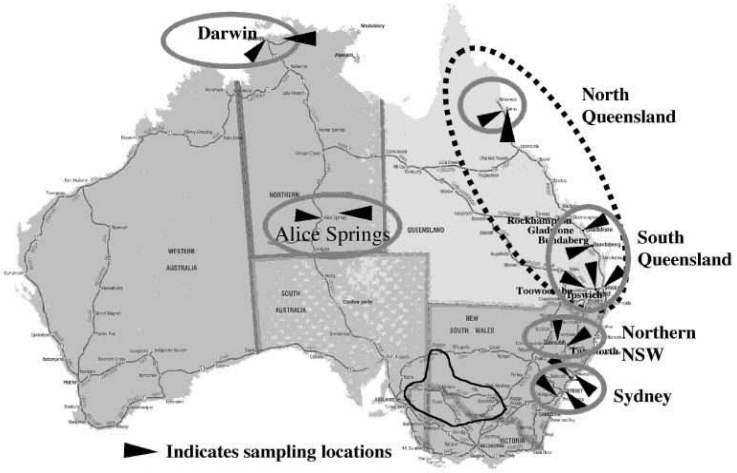

FIGURE 1. Sampling sites for the Australia-wide study

2.2. **The Australia-wide distribution.**

The data for six microsatellite loci were produced by Hong Yu [16]. She worked intensively for many months, analysing nearly 25,000 samples. This represented an average of over 40 flies at six loci for each of around twenty sampling sites for each of the five years of the study. There were remarkably few missing values in the data set.

Although initially we had little idea of what sort of population distribution to expect, it soon became clear that some of the samples close to each could be grouped as coming from the same population. I'm a

great believer in the old-fashioned χ^2 test statistic (see below), rather than in some of the fancier new genetic distance statistics. I wrote a small computer program to calculate χ^2 for any combination of samples tested against any other. Repeated use of this showed that the samples could be grouped into six regions as in Figure 1.

The most striking result, however, was the finding that the samples from North and South Queensland were completely homogeneous and could themselves be grouped. As is clear from the map, these regions are a long way from each other. But this is the home range of Qfly, and it seems clear that there is little if any population differentiation within this range.

Outside of this range there is clear differentiation between samples. Four regions outside of Queensland were identified, as seen in the figure. More on this below.

2.3. **Variation between years.**

The homogeneity of samples from different regions also extended to homogeneity of most samples from different years. In fact the one exception to this occurred in the sample in Figure 1 labeled 'Northern NSW', where the χ^2 values were quite significant.

A diagrammatic representation of the distribution is given in Figure 2. This is a 2-dimensional representation of the principal coordinates of the microsatellite frequencies. The diagram has nothing to do with actual physical distances. It simply reduces a set of frequencies of nearly 50 alleles at the six loci to the two dimensions that best summarise the variability in the data.

Clearly the Darwin and Alice Springs samples are well differentiated from the East coast samples. However the important point shown by the figure is that the points remain in the same area for each of the populations except for Northern NSW. The diagram does not indicate significance, since the variability in this figure depends on the sample size as well as the allele frequencies. Queensland, for example, is the most stable because of the large sample size.

What seems to have happened in the Northern NSW population is that it was well differentiated from Queensland, and to a lesser extent from Sydney, for the years 1994 - 1996. Then in the years 1997 and 1998 the samples became indistinguishable from the Queensland samples. Presumably there was some sort of invasion from the much larger Queensland population in those years. However it is possible that the

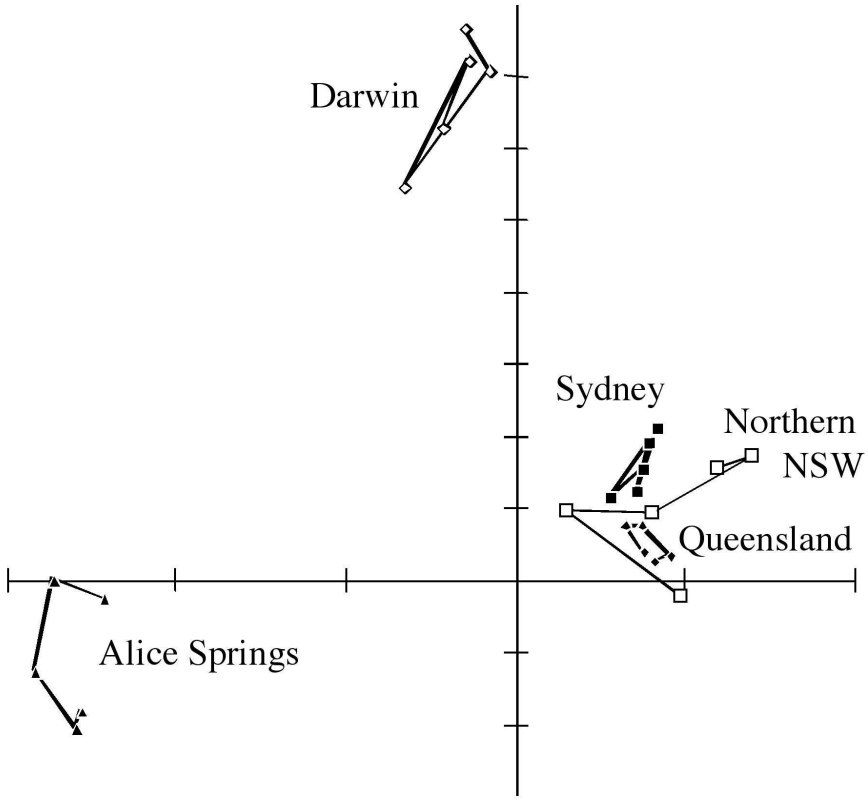

FIGURE 2. 2D analysis of microsatellite frequencies for different years in the five regions

previous three years were anomalous, since studies since have shown a very low level of differentiation between Queensland and Northern NSW samples.

2.4. **An aside on the use of the χ^2 test.**

I was raised, statistically speaking, in a department heavily influenced by RA Fisher, often with lectures from the great man himself. Unlike the analysis of variance and many other standard statistical techniques, Fisher didn't actually invent χ^2, which dated from before his time. He did, however, work out the basis for the heterogeneity or contingency test, the most common use of χ^2, and personally calculated the probability tables.

His use of χ^2 that I recall most vividly was the analysis of Mendel's data [3], in which he showed that Mendel's data agreed too closely with expectations. The idea that data can agree too closely with expectations may seem a strange one, but when one is doing counts for repeated random events, a certain level of chance deviation is always expected. Many things can happen to make the deviations greater than expected, but it is much more difficult to explain away the lack of deviation. Cynical teachers will often apply such a test to 'data' produced in undergraduate experiments to indicate that the student has invented the data rather than bothering to do the counts.

The χ^2 test shows the expected deviation very clearly. Where there are many classes being tested, and/or a number of tests are combined, the expected value of χ^2 is equal to the number of 'degrees of freedom'. So, for example, if 100 independent tests for agreement with a simple ratio are combined, giving 100 degrees of freedom, then a value of χ^2 significantly greater than 100 indicates some positive deviation from the hypothesis, whereas a value significantly less than 100 indicates something 'fishy'. Fisher used to claim that it was almost impossible for someone to invent a large data set that a competent statistician couldn't show had been artificially manipulated. He, incidentally, showed not only that Mendel's data agreed too closely with expectations, but that in one case Mendel got the expectations wrong and the data agreed too closely with the incorrect expectations but departed from the correct expectations. Fisher liked to think that Mendel's gardener knew what was expected in the experiments and threw out some of the cases where deviations looked too large.

Anyway all of this is preliminary to looking at Hong's data set for the microsatellites from the Double Helix study. With nearly 25,000 observations, there were many possible χ^2 tests, including heterogeneity within regions, between regions and between years. I'll show, for example, one test that indicates that Northern and Southern Queensland are part of one large population.

Year	1994	1995	1996	1997	1998	Total
χ^2	67.7	73.6	53.2	62.7	68.4	325.6
d.f.	76	58	60	58	66	318
N	2316	2256	1980	2100	2100	10752

In all cases the χ^2 values are appropriately close to the number of degrees of freedom. The sample size in each case is shown in the bottom row.

Other tables are shown in the publication Yu et al (2001) [16]. The values show a satisfyingly consistent picture of isolated populations stable over years, save for the one example where invasion has occurred from the large nearby core population. Overall it is the cleanest large data set whose analysis I have been involved in.

2.5. **Implications for control of Qfly.**

The most important conclusion coming from the study is that there is a core population of Qfly in Queensland, and genetically differentiated populations outside of this range. This genetic differentiation may, or may not, indicate adaptation, as discussed below.

From the point of view of control of Qfly, the important populations are those on the East coast of Australia. This control is not necessarily relevant for the Northern NSW and Sydney regions, where the fly is endemic, and where little can be done at present. Where the control is important is in agricultural regions to the South and inland, particularly in the Fruit Fly Exclusion Zone (FFEZ) shown roughly in outline on the map in Figure 1, where occasional outbreaks occur.

The fact that the endemic NSW populations are so stable and differentiated from the main Queensland population strongly suggested to us that it is these populations, rather than the main Queensland population, that provide the source for outbreaks. Stuart Gilchrist and I [12] analysed this situation further using a more detailed set of populations in the region of the FFEZ. We noted that some of these populations shared alleles that were comparatively rare in the Queensland population, and developed a test for the use of such alleles to show that there must be intermediate population. These conclusions were also supported by heterozygosity calculations, showing where variability has been lost in the founding of new populations.

More recent studies by Stuart Gilchrist using an extended set of microsatellites and analysing samples on a much smaller scale from particular towns and outbreak samples have confirmed these conclusions. It is now getting close to the situation where the source of outbreaks can be pin-pointed to resident populations usually associated with particular towns.

Efforts to control the movement of fruit should evidently concentrate on local movements rather than long distance movements from the North. These policies seem now to have been accepted by the relevant agriculture departments.

The differentiation between Queensland and the rest of the country suggests an interesting possibility. A long-term campaign for the control of Qfly could possible be based on the complete eradication of flies South of the main Queensland population. If these populations are, indeed, adapted to the more marginal climates, then it is possible that the flies might not easily be able to re-invade. Given that the invasion happened once already around 100 years ago, this may seem a risky proposition. However the original invasion may have occurred over a reasonably long period required for genes for adaptation to be selected.

One area where evolutionary theory is on our side is the likelihood that such genes for adaptation, should they have been selected, are unlikely to have spread to the core Queensland population. The relative sizes of the populations, and the adaptation to the original climate regimes, make it likely that the core population has been unaffected by the southern invasions.

The chance that complete eradication in southern regions will be attempted, or would be successful, is, unfortunately, low. Total eradication of Medfly has been achieved by Chile, and by some states in Mexico. My colleague Alfie Meats tells me, however, that it would be more difficult to achieve in Qfly. And it is unlikely that the political will exists to even seriously contemplate such an exercise.

2.6. **Northern Territory populations.**

The Darwin and Alice Springs populations of Figure 1 are both more genetically differentiated from the Queensland population. The origin of the Alice Springs population seems clearly to have been a propagule of around ten flies. Emilie Cameron, who did a PhD on these populations, showed that the Alice Springs population has also led to at least one further invasion, of a small agricultural area called Ti-tree.

Emilie analysed much more widely in Northern and Western regions, and found that flies closely related to the Darwin population occur throughout the region. The origin of these flies, whether by straight invasion or by some sort of hybridisation event, is not really clear.

2.7. **The *B. tryoni* complex.**

B. tryoni has a closely related sibling species, *B. neohumeralis*. The two species coexist in areas of Queensland, although the range of the latter is much smaller. A third species, *B. aquilonis*, was reported to

occur in the Northern Territory. However following Emilie Cameron's study, the status of this species is in some doubt.

There is no similar doubt about *B. neohumeralis*, however. The co-existence of these two species is a matter of considerable interest, and is considered briefly in the next section.

3. THE GENETICS OF QFLY

Our original intention in the Qfly project was to develop a set of visible and molecular markers as the basis for a 'genomics' study. We have had mixed success in this.

3.1. **Visible markers.**

Our main approach to find useful visible markers was based on in-breeding. This approach has apparently been successful in a variety of organisms, although it is difficult to find references for this. My authority on this was Mel Green, Professor Emeritus at Davis, who used to visit Australia regularly and undoubtedly knew more than almost anybody else post HJ Muller about mutation.

In Drosophila it is easy to introduce wild strains. Just getting a single female from the wild and putting her in a tube of food will almost always lead to a line of flies. Unfortunately this rarely works in Qfly.

We developed a compromise technique in which wild males were introduced to laboratory-adapted females. This produces offspring in a reasonable fraction of cases. Single pairs can then be crossed, and their offspring intercrossed to produce sib-mating. In theory if an the original wild type male had a single recessive mutation there should be a reasonable chance that this will show up in recognisable homozygous form.

Sheelagh Cuneen and Pat Maheswaran, both I'm fairly sure astute observers, stared at thousands of F1 and F2 progeny. For whatever reason, the results were meagre. Some eye colour mutations were found. These all turned out to be allelic to each other, and the locus designated as *orange eyes* (*oe*) -probably homologous to Drosophila *scarlet*. But we never found a white-eyed mutation, although these are known in many insects and reputed also to have been seen in Qfly.

Two other mutations were found, *white marks* and *bent wings*. These were both useful in our attempts to produce a strain for 'male only

sterile release' (see below). Much effort went into screening for bristle markers, many of which are known in Drosophila, but none was found.

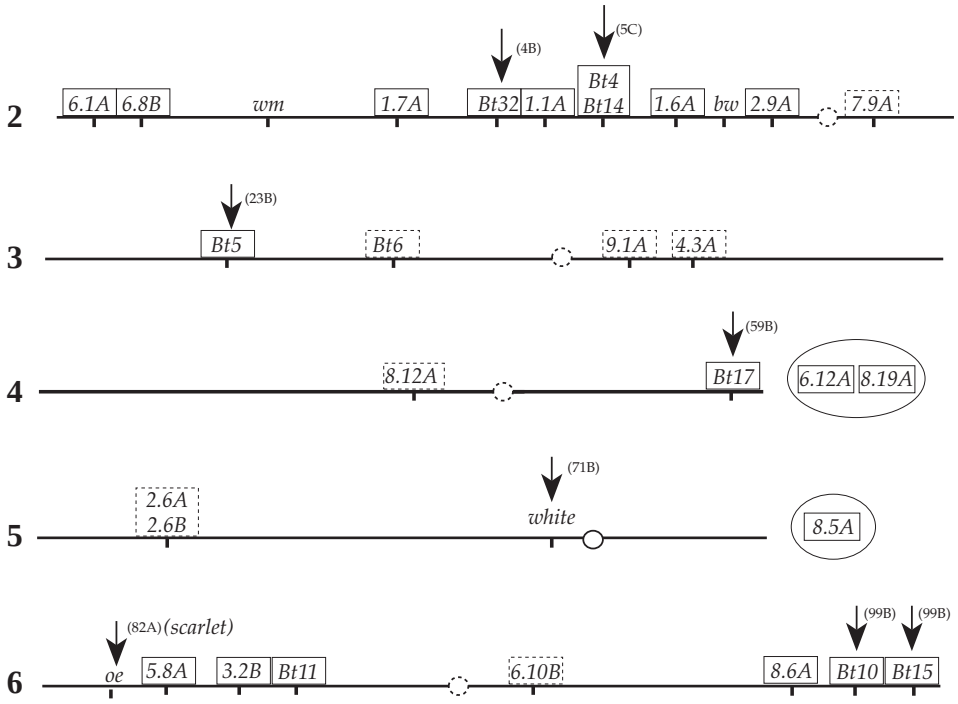

FIGURE 3. Map of microsatellite, visible and molecular markers

3.2. **A polytene chromosome map for Qfly.**

One of the most useful tools in Drosophila has been the detailed chromosome map produced from the giant chromosomes of salivary cells. Such chromosomes are usually available in other insects, although often difficult to prepare. Such turned out also to be the case in Qfly. However after much effort, literally years of effort, JingTing Zhao in our lab learned to prepare and recognise individual chromosomes, and to use *in situ* hybridisation to map individual markers. She did this under the tutelage of Antigone Zacharopoulou who similarly spent years working out how to do this in Medfly. JingTing and Antigone were able to recognise similarities in banding patters in order to use the same numbering of the five autosomes as had been used for Medfly.

The result of all this is the map shown in Figure 3. This includes the original six microsatellite markers, an additional 20 or so microsatellite markers from a later screen, the three identified visible markers and

two molecular markers. Some of this mapping was done using *in situ* hybridisation to particular bands in the polytene chromosomes, shown by arrows on the map. Most was done by standard mapping in crosses, the analysis of which was my small contribution to the effort. For the genetically mapped markers, the positions on the polytene map are only very approximate.

JingTing, who did the chromosome work, has now left the lab. The chances of anyone else learning to do this are very low. It's an example of what is becoming rarer in science, a unique personal skill that can't easily be transferred from one person to another. I once had a vision of working up a computer system to document banding patterns and identify individual chromosomes, but this was probably pie in the sky. This is a real problem if we ever need to extend the mapping and we can't call on JingTing.

3.3. **The sterile insect technique (SIT).**

The only non-chemical method currently available for control of a pest such as Qfly is through the release of large numbers of sterilised flies. The sterile males mate find wild type females, mate with them, and thereby sterilise them, so the theory goes. There is evidence that SIT works in some cases, the best-known example being screw-worm. SIT is widely used for Medfly. For Qfly, there is a small 'factory' on the outskirts of Sydney where flies are bred, of the order of a few million a week at peak time, the flies are sterilised at a nearby nuclear facility, and then released as pupae.

A problem with SIT in its most basic form is that large scale breeding produces large numbers of females as well as males. Extra irradiation doses needs to be given to ensure that fertile females are not released, and the evidence also indicates that the sterile males are less likely to look for wild females if many sterile ones are available at the point of release.

With millions of flies being produced, sexing on an individual scale is impossible. However there are ways of producing so-called genetic sexing strains. In Medfly the Y chromosome, which contains the genes causing the fly to be male, has been joined on to another chromosome containing a particular wild type gene. A mutant form of this gene causes lethality in double dose, but in males, if the mutant gene is protected by the wild type gene, flies are normal. This system needs to be 'conditional', otherwise all females would die. In Medfly the lethal gene is temperature sensitive, so that a burst of high temperature at the

early larval stage kills the females. For an account of the procedures see eg. http://www.ars.usda.gov/is/AR/archive/jun00/flies0600.htm.

We attempted to produce a similar strain in Qfly. A suitable mutation in our case was the *bent wings* mutation, which is unable to fly, and therefore effectively unfit in the wild. Furthermore we found that the mutation was sensitive to developmental temperature and therefore potentially a fully-fledged temperature sensitive lethal. Therefore our aim was to translocate the Y chromosome onto the chromosome containing this mutant gene, chromosome 2 (Figure 3) to produce a genetic sexing strain. We irradiated the flies and were able to select four cases of Y-2 translocations, the joining that we required. Unfortunately there was so much sterility in the strains produced that we were forced to abandon the exercise. The saga was written up in [8].

Other work in the laboratory has been undertaken to improve the performance of flies released from the factory. The most promising line of work involves hybridisation between established factory strains and wild type flies to try to nullify some of the weaknesses selected when flies become adapted to factory conditions. This work, undertaken by Stuart Gilchrist, is currently under way.

3.4. **Genetic differences between *B. tryoni* and *B. neohumeralis*.**

Jennifer Morrow [9] sequenced mitochondrial and *white* DNA in a number of individuals. Surprisingly few differences were found between these two apparently well established species. Even at the level of microsatellite frequencies there are few differences [15].

A major isolating factor in the field is time of mating - *B. tryoni* mates at dusk and *B. neohumeralis* mates during the day. When placed together in the laboratory, however, the two species cross reasonably easily. Hybrids are fertile, can be backcrossed to both parents etc. The co-existence of this species pair is one of the best documented cases of its kind, and work continues to try to understand the genetic and physical factors involved.

4. MOLECULAR STUDIES IN QFLY

Marianne supervised many Honours and PhD students in different areas that I won't describe here since I had little or no involvement. This involved isolating and characterising genes controlling sex determination, eye colour genes, timing genes, transposons etc. Some details

of the various projects and students involved can be found at the old FFRC website https://qffrl.wordpress.com.

4.1. **Genome sequencing.**

Genome sequencing started to become affordable not too long into the 21st century. Around 2011 - 2012 we decided that maybe it was time to think about a genome sequence for Qfly. The FFRC moved to University of NSW from Sydney University at that time, partly because of the Ramaciotti Centre for genome sequencing at UNSW.

One of the known difficulties of genome sequencing is to have a strain without too much variability. We chose the Bent wings strain that had come from a single mutation and had been maintained in small numbers in the lab for many generations, which therefore should be reasonably homozygous. Later in the project the technology had improved sufficiently that single flies could be used.

Deb Shearman made most of the DNA. It was sent to the Ramaciotti Centre Illumina machines. Out came millions of 100+ base pair sequences that Stuart Gilchrist put together to produce a genome sufficiently complete to get published [5]. This genome contained many thousands of 'scaffolds', DNA sequences ranging in size from a few Mb down to very small sequences. It served as the basis for a long time for the subsequent analysis, although now has been superseded.

4.2. **Genome mapping.**

My contribution came mainly from an experiment to map the scaffolds onto chromosomes. Genetic mapping uses parents and offspring genotypes to infer linkage relationships. Our experiment used three crosses between a single female and a single male, each cross producing around 30 offspring. The aim was to order the scaffolds, at least the larger ones, into a genetic map. Crosses between single female and single female are not easy in Qfly, and Stuart Gilchrist made many such crosses in order to get three with sufficient progeny.

We used a technique known as Genotyping by Sequencing (GBS). The technique picks out small random sequences scattered over the genome that happen to have two copies of a particular restriction enzyme site close to each other. Most large scaffolds will have at least one of these GBS sequences. Furthermore, with enough variability between sequences it is usually possible to tell which parent the particular scaffold in a given offspring comes from, the basis of genetic mapping. Rather than using the Bent wings stock that was used to minimise

variability for the genome sequencing, we used a stock labelled S06 that was descended from a collection made in Sydney in 2006 that was expected to contain more variability.

We didn't do the experiment ourselves. Rather we, actually my colleague Yizhou Chen from NSW Ag, made a 96 well plate containing the DNA from each parent and offspring of our three crosses. This was posted off to a group at Cornell University who specialised in GBS. They used the particular restriction enzyme to cut the DNA, then added small DNA sequences to define which of the 96 wells the sample came from, mixed the samples, and fed the DNA into an Illumina machine that sequenced the first 100bp of these sequences, A file containing approximately 450,000,000 of these 100bp sequences was then emailed back to us. This is an oversimplified account of what actually happened. We were very impressed by the Cornell group, and saddened to hear some time later that they were forced to stop the GBS work allegedly due to a patent infringement. Fortunately other variants of the GBS technique are still available commercially.

I wrote a series of around 15 Python scripts to analyse the data and come up with a genetic map for the scaffolds. This is a 'pipeline' where the results from one step are fed into the next program, so that each program does just one task. It's too complicated to describe here. I wrote a description of what the various programs do, mainly for my own benefit but available here. Copies of the Python scripts are also available at:
www.handsongenetics.com/PIFFLE/Pythonmappingprograms
although they are almost entirely undocumented and unlikely to be useful for any data set other than the one they are designed for.

One feature of mapping in insects is the absence of recombination in the male. This immediately simplifies the process, since markers (scaffolds defined by GBS sequences) on the same chromosome will always segregate together in the male. With a few exceptions to do with sex chromosomes mentioned below, the progeny in all three crosses fell into five classes when the segregating marker came from the male parents. The five classes represent the five major chromosomes. This immediately simplifies the analysis to ordering markers within the five chromosomes using recombination in the female.

The basic method of mapping using female recombination is simple. Say one has three markers, a, b and c. One of the parents is a b c / + + +. The highest class of offspring genotypes will be a b c and + + +. The next highest classes are, say, a + c and + b +. Then

one assumes that the correct order can't be a b c, ie b in the middle, since this would require that the highest recombinant class is a double recombinant. (Note that c b a is the same as a b c since direction doesn't come into the argument unless there is some external fixed marker). If a and c always go together then the only things one can say is that the order is either a c b or alternatively c a b. if the class a + + or + b c turns up, then it is clear that the correct order must be a c b.

For a large number (thousands) or markers, one first checks which ones are never separated by recombination. The limited number of progeny, or limited number of crossovers, means that this type of analysis cannot order close markers. Fortunately, however, these are the ones most easily ordered by sequencing.

I've never seen a description of how to extend the above argument for 3 markers to the case of thousands of markers. What the program in my set does is to try different orders to see which order minimises the number of crossovers needed to produce the observed offspring. The number of potential orders is, of course, huge, so that the difficulty is to somehow hone in on the order needing minimal crossovers. I've tested the program out on simulated offspring genotypes. Perhaps surprisingly, it almost always found the correct order.

One problem with the procedure is that each of the three crosses produces a different result, with some, but by no means complete, overlap of markers. I've attempted to combine the three maps into a single best map. Arbitrary decisions on gene order have to be made for markers mapped in different crosses, so the overall map seems to be of less use than the three individual maps.

A new genome has now been constructed by the group led by John Oakeshott in ex-Entomology CSIRO Canberra. Rather than hundreds of disconnected scaffolds, the new genome contains single sequences covering large portions of the genome. It has been made using a newer technique called HiC sequencing that allow adjacent regions to be joined together. I'm told that there is a chance of false joining, but that the availability of the mapping data reduces the chance of such errors.

4.3. **Matching genetic maps with polytene chromosomes.**

We have had a physical chromosome map for a long time (Figure 3). When the first mapping analysis separated scaffolds into five groups, it was reasonably easy to show which of these groups corresponded to

which of the five polytene chromosomes. The microsatellite markers were important here, but other markers such as *white* and *orange eyes* that had been assigned chromosome positions could also be used.

As mentioned above, genetic mapping leads to an arbitrary orientation. The obvious way of orienting the genetic map, as has long been done in Drosophila, is to use the direction implied by the polytene map. The Drosophila map has of course taken this process much further such that each sequenced region corresponds to particular polytene bands.

The orientation of the genetic map has been quite easy for two of the chromosomes. Chromosome 6, for example, could easily be oriented by the placement of microsatellites Bt10 and Bt15 at one end since the sequences of both these appear in different scaffolds. However even now (October 2020) we have no information on the orientation of chromosomes 3 and 5. Many sequences are available from all regions of the genetic maps, but the *in sutu* hybridisation experiments are yet to be done.

4.4. **The comparison with Drosophila.**

An important part of what we were trying to do is to compare the Qfly map with the established *D. melanogaster* map. This can't be done by direct comparison at the DNA level, owing to the high level of differentiation. However comparison of genes at the protein level is straightforward even for species separated by an estimated 60-70 mYr. Around 2,000 Qfly genes were located on mapped scaffolds and their Drosophila homologues identified. In fact the process is really the other way around since usually we initially identify Qfly genes by their Drosophila homologues. The result of all this is the map shown in Figure 4 (see [13]).

I'm rather proud of this map, all based on the results from one 96-well plate. There are five major elements in insect species called the Muller elements, and a small sixth one (see below). The map shows clearly the homologies with the five elements in *D. melanogaster*, with only around 10% of markers translocated from their original chromosome. The orientations within Qfly chromosomes are arbitrary, But there have evidently been so many inversions within elements between the two species that whichever orientation is chosen makes little difference.

I have some guilt feelings that our paper [13]) made little mention of the results from the Drosophila - mosquito comparison, a much more distant comparison of an estimated 250 mYr. I know little of the

mosquito work, and the comparison between genomes is more complex in this case. But the same basic elements seem to be distinguishable, in this case all identified using the much more labour intensive *in situ* hybridisation technique. I imagine that anyone from the mosquito world would not be impressed by our conclusion regarding chromosome stability from the comparison between two much more closely related species.

This preservation of the Muller elements over literally billions of generations is a very striking aspect of insect evolution. Most other orders change their basic chromosome structure at a much higher rate. In writing the paper, I was struck by a report that although the enzyme telomerase is present in many organisms, it is absent in insects. Insects maintain the ends of their chromosomes using a repeat sequence mechanism. I thought it worth speculating that the chromosome stability is due to this absence of telomerase. One of the reviewers objected to the inclusion of this argument. I have to admit that it is a speculation that is probably untestable.

4.5. **X linkage.**

When we made the DNA from individual flies, we were concerned about autosomal mapping and paid little attention to sexes. In fact I even managed to lose the listing of which DNA came from which sex on the plate. Fortunately it was quite easy to infer which parent was female and which male based on recombination in the offspring. And then two scaffolds defined by GBS markers segregated independently from the five mapped chromosomes, giving patterns strongly suggesting X linkage, which allowed us to unequivocally sex the offspring.

Just around the time that we were getting our results, I found out about a paper by Vicoso and Bactrog [14] claiming the small dot chromosome in *D. melanogaster* was descended from an ancestor of the X chromosome in a Bactrocera fruit fly, *B. oleae* and other non-Drosophilid species. They identified many Bactrocera X chromosome genes simply by using excess of female to male DNA reads. Deb Shearman had previously suggested that the missing small sixth Muller element might be contained on the X chromosome, which I had found hard to accept based on our failure to find many such X-linked genes. Subsequently it became clear that this failure was due to the low variability on the Qfly X chromosome, and that maybe we were lucky to find the two X-linked scaffolds by GBS that we did. As predicted by Vicoso and Bactrog [14], we found using their criterion that the majority of Qfly genes with *D. melanogaster* chromosome 4 homology were X-linked.

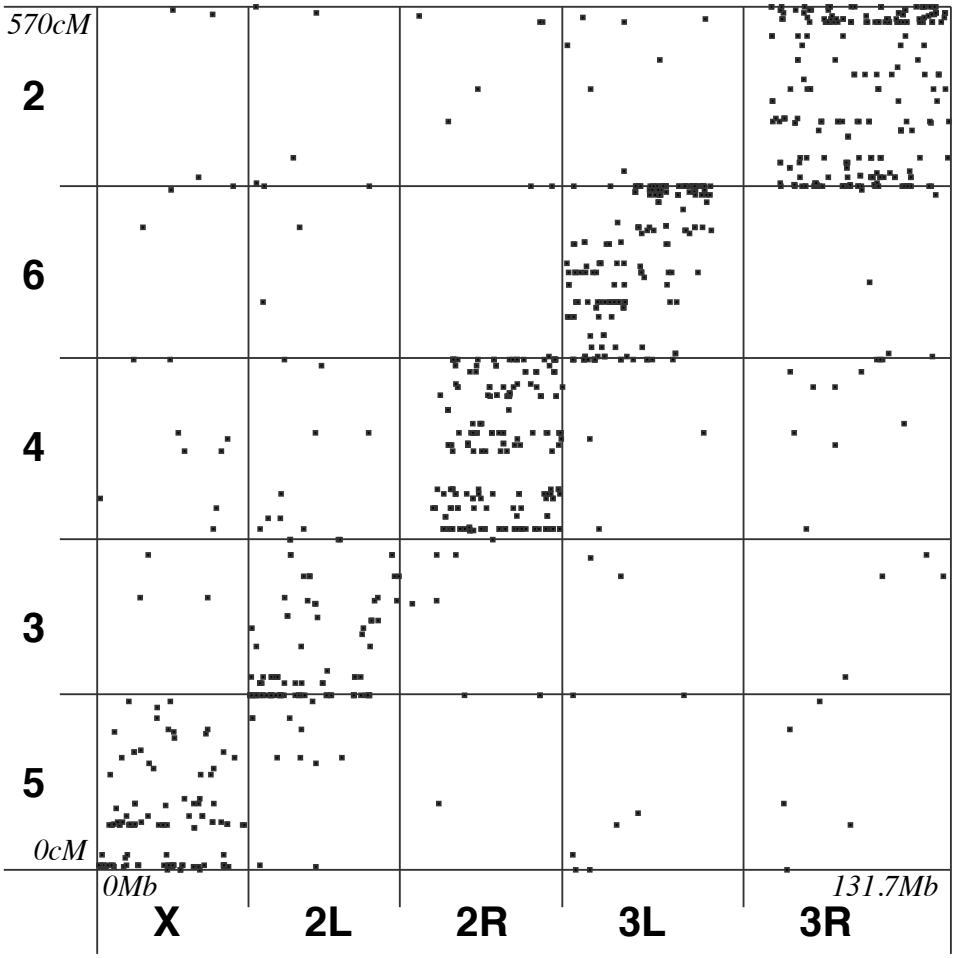

FIGURE 4. Comprison of homologous genes in Drosophila (horizontal axis) and Qfly (vertical axis)

This change of function of a chromosome from one that determines sex to just a small autosome is a striking finding, but consistent with a lot of evidence that sex determination is a very malleable process.

4.6. **Y linkage.**

In a further development, we were able to identify potential Y-linked scaffolds using the excess of GBS reads in males. In theory there should be many reads in male offspring and zero in females. Many cases of this type were found, but the whole exercise turned out to be more complicated, due to closely related homologues on the autosomes to most Y chromosome regions.

Amanda Choo and Simon Baxter from University of Adelaide were colleagues on this work. Part of the reason was to identify Y chromosome regions that could be used to insert genes using Crispr. It was important to identify regions that lacked homology with any autosomal regions. A dozen or so scaffolds were identified and matched to bulk male and female DNA. A combination of using both exact and weak homology, and selecting for regions that definitely possessed the two GBS restriction enzyme sites, allowed a few potentially legitimate Y regions to be identified. Amanda designed PCR primers and showed that they correctly hybridised only to male DNA.

Simon looked more closely at some of the identified Y chromosome scaffolds, and showed that one contains a unique duplicated gene that is common to Y chromosomes of many Bactrocera species [1]. Although its function is currently unknown, its conservation over evolutionary time shows that it must play some role.

4.7. **The Lewontin-Birch hypothesis?**

When we obtained the first genome sequences from *B. tryoni* and *B neohumeralis*, one observation stood out for me. A particular *B. tryoni* scaffold of length around 400kb matched smaller *B neohumeralis* sequences almost exactly over most of its length. Most other *tryneo* comparisons showed considerably higher genetic distances, as one would expect from two species.

What this immediately brought to mind was a hypothesis suggested by the well known evolutionary biologist Richard Lewontin early in his career when he visited Charles Birch's lab at Sydney University [7]. As explained in Section 2.5, Qfly in believed to have invaded from Queensland into NSW and more southern areas within the past 200 years. Lewontin proposed that the strain that invaded was a hybrid of *B. tryoni* and *B. neohumeralis*, and that it was this hybridisation event that allowed the range expansion. There was no direct evidence for this hypothesis. Now, however, it seemed that we had direct evidence. The Qfly strain that we used was of Southern origin, and it appeared to contain *B. neohumeralis* DNA. I gave a talk on this at the Genetics Society of Australasia meeting, and excitedly wrote to Lewontin to tell him that his hypothesis appeared to be confirmed.

A year later I gave a poster entitled "Apparent refutation of the Lewontin-Birch introgression hypothesis in Queensland fruit fly" to the GSA meeting. What had changed? The answer was that we realised that all results involving *neo* had been obtained using a single lab strain,

*neo*B. This had the day mating and callus colour expected of a *neo* strain. However we managed to get some wild type *neo* flies, and in the comparison of DNA from these flies and the *try* scaffolds the close homology had disappeared.

What must have happened is that instead of looking at an introgression event in wild southern *try* flies, we were looking at an introgression event in our lab *neo*B strain. Figure 5 shows what seems to have happened. The consequences of the two possibilities are the same if you don't know which is *try* DNA and which *neo* DNA for this region.

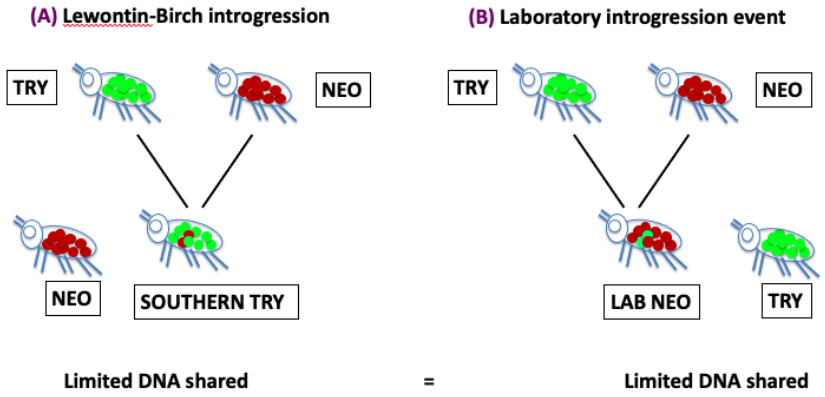

FIGURE 5. Alternatiive possibilities

One reason I was reluctant to accept the second explanation is that *try*-*neo* hybridisation is not supposed to go this way. Hybrids between the two species can be produced, but these individuals tend to have dusk mating characteristic of *B. tryoni*. The supposition, therefore, is that any hybrids will tend to be absorbed into the *B. tryoni* pool as hypothesised by Lewontin. Evidently the opposite must have happened here, so that somehow a portion of *try* DNA has been introgressed into the *neo* pool, in spite of the fact that *try* mating behaviour is dominant.

Following mapping of the chromosomes, it became clear that scaffolds with a high *try*-*neo* homology were concentrated at one end of chromosome 5. There are one or two other possible outliers, but it seems likely that just this region of the *try* genome has been introgressed into our lab *neo*B strain.

Before we realised that this introgression had occurred, this *neo*B strain was used in several experiments. Kathie Raphael's extensive studies on *try* - *neo* expression differences [11] used this strain. Fortunately the introgression appears to have no effect on the timing of mating, so the introgression is unlikely to have had any effects on the conclusions. Similarly the strain was used by colleagues at CSIRO in experiments on qtl mapping of behavioural differences. Here also the conclusions seem robust and unaffected by the introgression.

References

[1] Amanda Choo, Thu N M Nguyen, Christopher M Ward, Isabel Y Chen, John Sved, Deborah Shearman, Anthony S Gilchrist, Peter Crisp, and Simon W Baxter. Identification of y-chromosome scaffolds of the queensland fruit fly reveals a duplicated gyf gene paralogue common to many bactrocera pest species. *Insect Mol Biol*, 28(6):873–886, 12 2019.

[2] R.A.I. Drew. The tropical fruit flies (diptera: Tephritidae: Dacinae) of the australasian and oceanian regions. *Memoirs of the Queensland Museum.*, 26:1–521, 1989.

[3] RA Fisher. Has mendel's work been rediscovered? *Annals of Science*, 1:115–137, 1936.

[4] M Frommer, L E McDonald, D S Millar, C M Collis, F Watt, G W Grigg, P L Molloy, and C L Paul. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual dna strands. *Proc Natl Acad Sci U S A*, 89(5):1827–1831, 1992 Mar 1.

[5] Anthony Stuart Gilchrist, Deborah C A Shearman, Marianne Frommer, Kathryn A Raphael, Nandan P Deshpande, Marc R Wilkins, William B Sherwin, and John A Sved. The draft genome of the pest tephritid fruit fly bactrocera tryoni: resources for the genomic analysis of hybridising species. *BMC Genomics*, 15:1153, Dec 2014.

[6] M. W. Kinnear, H. S. Bariana, J. A. Sved, and M. Frommer. Polymorphic microsatellite markers for population analysis of a tephritid pest species, bactrocera tryoni. *Mol Ecol*, 7(11), 1998.

[7] R C Lewontin and L C Birch. Hybridization as a source of variation for adaptation to new environments. *Evolution*, 20(3):315–336, Sep 1966.

[8] AW Meats, P Maheswaran, M Frommer, and JA Sved. Towards a male-only release system for sit with the queensland fruit fly, bactrocera tryoni, using a genetic sexing strain with a temperature-sensitive lethal mutation. *Genetica*, 116:97–106, 2002.

[9] J Morrow, L Scott, B Congdon, D Yeates, M Frommer, and J Sved. Close genetic similarity between two sympatric species of tephritid fruit fly reproductively isolated by mating time. *Evolution*, 54(3):899–910, 2000 Jun.

[10] R Osborne, A Meats, M Frommer, JA Sved, RAI Drew, and MK Robson. Australian distribution of 17 species of fruit flies (diptera: Tephritidae) caught in cue lure traps in february 1994. *Australian Journal of Entomology*, 36:45–50, 1997.

[11] K A Raphael, J A Sved, S Pearce, J G Oakeshott, A S Gilchrist, W B Sherwin, and M Frommer. Differences in gene regulation in a tephritid model of prezygotic reproductive isolation. *Insect Mol Biol*, 28(5):689–702, 10 2019.

[12] J. A. Sved, H. Yu, B. Dominiak, and A. S. Gilchrist. Inferring modes of colonization for pest species using heterozygosity comparisons and a shared-allele test. *Genetics*, 163(2), 2003.

[13] John A Sved, Yizhou Chen, Deborah Shearman, Marianne Frommer, A Stuart Gilchrist, and William B Sherwin. Extraordinary conservation of entire chromosomes in insects over long evolutionary periods. *Evolution*, 70(1):229–34, Jan 2016.

[14] Beatriz Vicoso and Doris Bachtrog. Reversal of an ancient sex chromosome to an autosome in drosophila. *Nature*, 499(7458):332–5, Jul 2013.

[15] Y Wang, H Yu, K Raphael, and A S Gilchrist. Genetic delineation of sibling species of the pest fruit fly bactocera (diptera: Tephritidae) using microsatellites. *Bull Entomol Res*, 93:351–360, 2003.

[16] H. Yu, M. Frommer, M. K. Robson, A. W. Meats, D. C. Shearman, and J. A. Sved. Microsatellite analysis of the queensland fruit fly bactrocera tryoni (diptera: Tephritidae) indicates spatial structuring: implications for population control. *Bull Entomol Res*, 91(2), 2001.