From molecules to mushrooms: DNA sequencing in Norfolk

Tony Leech, Mike Ball, Steve Judd, Gill Judd, Tony Moverley & Yvonne Mynett

The double-helical structure of DNA had been established just a decade before I (TL) began my biochemistry degree. It was an exciting time as further work soon explained how the DNA molecule could contain the information to create a living organism and, in broad terms, how this is implemented. But I never imagined that this knowledge would be used to identify species and illuminate our understanding of evolution. Even less that, sixty years on, I would be associating with amateur naturalists doing precisely this. This article describes how members of the Norfolk Fungus Study Group have been able to extract DNA from fungi, have the relevant parts of their base sequence determined and interpret the results.

Basics

Deoxyribonucleic acid (DNA) is a very large molecule but, in essence, a simple one. It consists of bases, chemically linked to form a long chain. The bases, themselves molecules, are of just four kinds, now universally designated by their initial letters: A, C, G and T. DNA provides the information for the synthesis of proteins which form the structures, directly or indirectly, of organisms, and determine how they function. This information is contained within the sequence, or order, of these bases so it follows that each species will have a unique sequence of bases in its DNA.

DNA normally exists as a double strand, one carrying this information and the second, complementary to it, bearing a sequence determined by the first but with A opposite T, and G opposite C. The prediction by James Watson and Frances Crick, discoverers of this double-helical structure, that this would provide a mechanism for exact copying was soon confirmed. Chemical bonds hold the bases on opposite chains together at normal temperatures but are sufficiently weak to be broken in the copying process or at high temperatures. When the chains are separated, complementary bases available in the cell align and are linked together by enzymes to form two identical double-helical molecules.

Determining the order of the bases along a DNA molecule, that is sequencing it, depends on some very ingenious chemistry. It took many scientists, 15 years and a phenomenal amount of money to determine the sequence of the 3.1 billion base-pairs of the DNA in a human cell nucleus (the genome) announced in 2003. Since then, many more species have had their whole genome sequenced as automation and even more ingenuity have seen enormous reductions in cost and increase in speed.

Each cell in a multicellular organism contains an identical set of DNA molecules formed by a precise DNA-copying mechanism when the cell divides. Actually, there is the potential for slight changes in base sequence as copying 'mistakes' occur at low frequency. These are the mutations which are one of the reasons why offspring of organisms reproducing sexually are not identical to their parents. If in a gamete (sex cell), a mutation occurs in a region of DNA responsible for determining a critical function, it is likely that the offspring will lack this function and fail to survive. However, there are stretches of DNA which either lack function or have functions which do not depend on the precise sequence of

bases. Mutations within these are much less likely to be lethal so will accumulate, albeit slowly, as generation succeeds generation. It is these regions, known as barcode regions, that mutate at a rate comparable with the evolution of new species which can be used for species identification. In fungi, one such region is the nuclear ribosomal internal transcribed spacer region (ITS) but others are being increasingly used in conjunction with ITS to clarify species separations.

Preparation of DNA for sampling

In order to sequence the ITS region of a fungal genome, multiple copies of that region, and that region alone, must be made. It is now possible for this to be carried out by non-specialists outside a laboratory.

After a very small amount of fungal tissue, preferably from the spore-producing region, is homogenised in an aqueous medium, primers for the ITS region are added. These are commercially-produced short sequences of DNA which are complementary to base sequences at each end of the ITS region. These act as markers to stop and start copying when polymerase enzyme is added (with bases). Once copying is complete, the old and new strands are separated by heating and a second cycle of copying initiated. Each cycle doubles the amount of DNA from the ITS region. As the polymerase used is not damaged by heating, the process is carried out in a thermocycler which is programmed to raise and lower the temperature alternately every few minutes (Fig.1). After an hour or so, many millions of copies of the targeted DNA will have been produced in this polymerase chain reaction (PCR).

To check that enough of an appropriately-sized DNA segment has been produced, electrophoresis is carried out on a sample. In this technique, molecules of DNA are induced to move through a gel at rates according to their size by the application of an electric field and can be visualised by staining. If the extraction has been successful, the sample is sent to a commercial laboratory for sequencing.

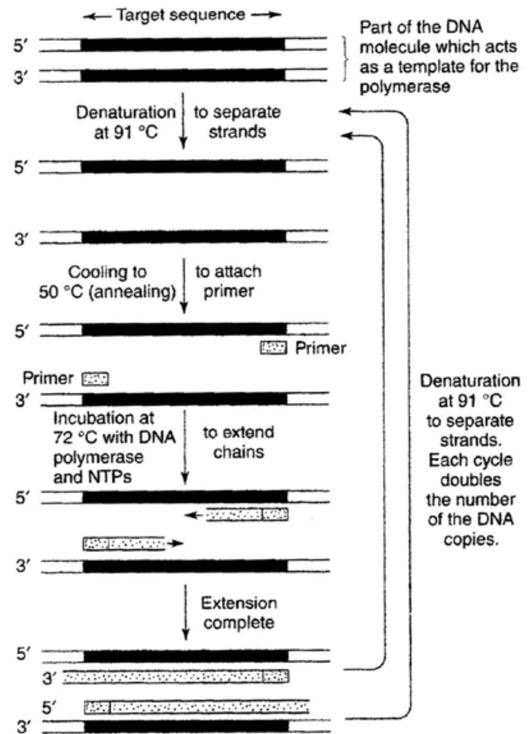

Figure 1. Polymerase chain reaction. The actual temperatures depend on the equipment and reagents used. Generally, only one chain (3′) is sequenced but both can be, adding to accuracy. *From: Functional Biochemistry, E.A. Newsholme & A.R. Leech. Wiley Blackwell (2010).*

Use of the ITS region in fungus identification

The use of DNA sequencing in species identification depends on the assumption that different species will have different base sequences in their barcode regions, but this assumption must be qualified. Even within the barcode region there are sequences which are not diagnostic and which must be removed to improve matching. Software programmes are available to facilitate this but their use involves some judgement and experience. The determined sequence is then compared with published sequences on one or more databases through a search

facility. This will typically throw up a range of possibilities with differing percentage matches.

By no means all fungal species have been 'barcoded', so getting a good match is not a certainty. But how good does a match have to be to be able to assign the specimen to a named species? Even within the barcode region, mutations will have occurred so that variants exist which are not recognised as separate species. This raises the question of what constitutes a species, a question that does not have a definitive answer. In practice, agreement in excess of around 98.5% is considered evidence of conspecificity, at least in the context of fungi and the ITS region..

There is, however, a fundamental weakness in the above procedure; it depends on the validity of the determination of the species which has contributed data to the database and with which the specimen is being compared. This is a particular problem with fungi which typically have a rather variable morphology and which have not been studied by taxonomists for as long as, and in as much depth as, plants and animals. It has been estimated that as many as 40% of the published sequences for fungi are based on erroneous identifications! By definition, this problem does not exist if the sequence of the holotype on which the species has been based has been sequenced but these are often not available, at least in a condition which allows sequencing. Judgement is needed in deciding whether to accept a match, often based on geographical information and whether the sequence has been determined by a fungal taxonomist.

Establishing a fungal DNA group in Norfolk

Although the preparation of fungal DNA for sequencing can be carried out by non-scientists, it does require some skills. The main one is in handling minute volumes of solutions accurately, sometime less than a two-hundredth of a millilitre – a twentieth of a small drop. Thermocycling machines and electrophoresis devices have been miniaturised and made user-friendly but their use, and the manipulation of micropipettes, requires some training and practice. DNA sample preparation typically takes a few hours but a number can be processed simultaneously.

Acquiring the equipment represents a one-off cost but reagents have to be purchased and there is a cost of sequencing. The cost per sample depends on a number of factors; the use of pre-prepared reagents is convenient but more expensive. Currently we estimate that it costs around £10 per sample but it has to be remembered that not all extractions are successful although those that are not do not incur sequencing costs.

In principle, DNA sample preparation can be done almost anywhere, including a kitchen table, but ideally the location needs to be easily accessible by those involved, have sufficient cleanable working space and allow storage (including a freezer and refrigerator) of reagents and equipment.

The Norfolk fungus DNA sequencing team has only become functional through the fortunate coming together of all these requirements. Preliminary investigations by TL of the possibility of working with Anne Edwards at the John Innes Centre to set up a sampling facility somewhere in Norfolk were going well until the Covid lockdown halted progress. Meanwhile, mycologists at the Royal Botanic Gardens Kew had developed good working relationships with the British Mycological Society (BMS) which supports both amateur and professional mycologists. Recognising the contribution that amateurs could play in research, a programme was developed to encourage amateur groups to become involved in DNA work and several have done so. As part of this programme, Brian Douglas, then employed at Kew, tutored workshops in Norfolk and facilitated a grant to provide equipment and materials. At much the same

time, the Darwin Tree of Life (DToL) project was being rolled out, supported by the Wellcome Foundation. The ambitious aim was to set in progress the determination of the full genome base sequences (not just the barcode regions) of all British organisms. The local 'hub' for this initiative was the Earlham Institute in Norwich. One of the ancilliary aims of DToL was to raise public awareness of the value of DNA sequencing and to enable schools and amateur groups to carry out DNA sampling. The local initiative was tagged 'Barcoding the Broads' and was organised by Sam Rowe who led further workshops and facilitated the obtaining of equipment and reagents.

A fortunate development at this time was the upgrading of workroom facilities at The Ted Ellis Trust, Wheatfen and the interest shown by the Norfolk & Norwich Naturalists' Society in locating its library there. The enthusiasm and cooperation of Warden Will Fitch has made it possible to set up a DNA sequencing laboratory at Wheatfen.

That leaves the workers. A small group (MB, SJ, GJ, TM & YM) with diverse skills and strong interest have met regularly to deliver the results described below. Not only have they mastered the manipulative and organisational skills required but also the necessary data-handling techniques. A further development has been the interest shown by mycologists from RBG Kew in using Wheatfen as a collecting site for fungi for whole-genome sequencing. Sample preservation for this work requires skills and precautions beyond those of amateurs. Visits have been made and further ones are planned.

Candidates for sequencing are proposed by members of the Norfolk Fungus Study Group and considered by the DNA team. Criteria include the confirmation of species new to Norfolk, especially those in genera in which it is difficult to resolve species using morphological features, and specimens which deviate in some way from published descriptions. Currently, most success has been had with agarics (gill fungi) but it is intended that other groups will be tackled.

Results

The first successful DNA extractions were carried out during training sessions late in 2021. Up to the end of January 2023, about 50 extractions had been attempted, most of which produced usable sequences (Table 1). At least some of the 'failures' were attributed to attempts on dried material which had been imperfectly stored

One of the fungi sequenced during training sessions, *Tarzetta alnicola*, was a new British record. An important caveat is that the absence of records on the national database (Fungus Record Database of Britain and Ireland (FRDBI)) does not mean that the fungus has not been recorded in Britain, only that the record has not found its way onto the database. There is always a delay between identifying and uploading a record and in the case of 'new' species it can be years before FRDBI can accept the record.

Tarzetta is a genus of medium-sized cup fungi, deeply cup-shaped and with pale colours. Two species, *T. catinus* and *T.*

Table 1. Summary of extractions to the end of January 2023. *The most likely reason for failing to find a satisfactory match is that a reliable reference sequence was not available.

Category	Number of extractions
Confirmation or identification of fungi new to Norfolk or otherwise	14 (including two repeats)
Confirmation or identification of more widespread fungi	22 (including two repeats)
No satisfactory matches*	6
Failed extractions	9

Figure 2. ***Tarzetta alnicola*****. Trinity Broad, Rollesby.**
Yvonne Mynett.

cupularis are widely recorded but difficult to distinguish and it is probable that much confusion has occurred. When YM found a cluster of small, pale *Tarzetta* on soil under Alder *Alnus glutinosa* at Trinity Broad, Rollesby in August 2022 (Fig. 2) she considered that they might be *T. alnicola*, a species newly described by van Vooren *et al.* (2019). This was confirmed with a 100% match with the holotype.

In 2017, YM identified the first British specimen of *Psathyrella thujina*, a small reedbed brittlestem, at Watermill Broad, Cranwich. This was subsequently confirmed by DNA sequencing at Kew. A year later, she found a second specimen at Hickling Broad which the DNA team was able to confirm, thereby adding a second British location (Fig 3).

Figure 3. ***Psathyrella thujina*****. Hickling Broad.**
Yvonne Mynett.

Entoloma ventricosum is a small pinkgill with grey-brown cap similar to more widespread species. The fungus was also collected by YM, this time at Alderfen Broad in April 2022 in association with Common Reed *Phragmites australis*. It appears to have been recorded from two sites in Scotland and two from North Wales (NBN Atlas).

Figure 4. ***Psathyrella tenuicula*****. Broadland Country Park.** *Mark Joy.*

At the end of December 2021, Mark Joy took some pellets of deer dung from Broadland Country Park for incubation. On them, a number of inkcap fruiting bodies appeared which YM was able to determine as *Coprinellus parvulus*, a species apparently unrecorded from Britain (Fig. 4). This identification was confirmed by national coprinoid expert, Derek Schafer. However, the DNA sequence registered a 100% match with *Psathyrella tenuicula*! It was then realised that these two 'species' had been synonymised and that the Norfolk record was, in fact, the second British record.

Figure 5. *Tulosesus (Coprinellus) velatopruinatus.* Bayfield. Top: mature. Bottom: young. *Tony Leech.*

Derek Schafer was also consulted when TL was shown a cluster of small inkcaps which had appeared in a flowerpot at Natural Surroundings, Bayfield (Fig. 5). TL's tentative identification was of *Coprinellus velatopruinatus*, a species hitherto only recorded from various Buckinghamshire sites and from RBG Kew, all by Derek Schafer. This identification was confirmed by the DNA team. In all probability, the species is much more widespread but requires informed microscopic investigation for identification. The fact that this species, with several others, has been transferred to the new genus *Tulosesus* (as *T. velatopruinatus*) underlines the widespread taxonomic reworking of these delicate dark-spored species which is being driven by sequencing information.

The genus *Cortinarius* (webcaps) has more British species than any other but until recently it was almost impossible to name many of them reliably due to conflicting literature descriptions and the plethora of synonyms. That changed with

Figure 6. *Cortinarius epipurrus.* **LIttle Plumstead.** *Tony Leech.*

a comprehensive study by Liimatainen *et al.* (2019) based on DNA sequences which was soon followed by a well-illustrated and accessible field guide to the group (Kibby, 2021) which has allowed mere mortals to make identifications. During October 2022, two species not appearing on the FRDBI database were collected by TL on public forays, *C. epipurrus* (Fig. 6) in a hedgerow under oak at Little Plumstead and *C. hirtus*

Figure7. *Cortinarius hirtus.* Briston. *Tony Leech.*

Table 2. Species confirmed through DNA sequencing that are first or second county records. All are gill fungi. *KR, Kerry Robinson.

Species	Date	Place	Recorder	Status
Amanita olivaceogrisea (Fig. 8)	3/10/2021	Wheatfen	KR* (TL)	2nd Norfolk; rare in UK
Coprinus sterquilinus Midden Inkcap	5/11/2022	Somerleyton	YM	1st Suffolk; scarce in UK.
Inocybe helobia	27/8/2020	Sutton Fen	YM	2nd Norfolk (at same site as 1st); rare in UK
Pluteus semibulbosus	20/9/2021	Potter Heigham	YM	2nd Norfolk; scarce in UK
Psathyrella lutensis	17/9/2022	Sutton Fen	YM	1st Norfolk; scarce in UK

(Fig. 7) under birch at Briston (Leech *et al*., 2023). Both were confirmed by sequencing and have been reported from elsewhere in Britain although details are yet to be published.

Figure 8. *Amanita olivaceogrisea*. Wheatfen *Tony Leech.*

Additional species of interest that have been confirmed by the DNA team in 2022 are listed in Table 2.

Future developments

It is envisaged that the DNA team will continue to sequence problematic and rare fungi. It was always the intention to involve others expressing a serious interest and commitment once procedures had been established. During 2023, Anne Crotty has joined the team.

It is hoped that the team will, in time, undertake research, probably in cooperation with a taxonomist. At least one amateur group has achieved this, namely the Pembrokeshire group under David Harries which has contributed to the description of new species of *Microglossum* (Harries *et al*., 2018).

In principle, the procedures established by the DNA team for fungi are applicable to other taxonomic groups but, at present, morphological characters are adequate for species identification for most insect orders and for vascular plants so there is no attraction in barcoding these groups. However, at least two interesting entomological discoveries have been made as a result of training workshops run as part of the 'Barcoding the Broads' initiative. In the first, plastic pipes in use at a Norfolk trout fishery as a substrate for fish lice to lay their eggs (and be removed) were found also to bear unidentified green eggs. Barcoding enabled these to be identified as a species of mayfly, the Pale Watery *Baetus fuscatus*

hitherto thought to be an exclusively running-water species (Collins, 2022).

The 'Barcoding the Broads' project has also introduced DNA sequencing to schools, including Wymondham College where sixth-form students extracted DNA from what was probably a planthopper nymph collected on the campus. However, the sequence very closely matched that from the big-headed fly *Pipunculus zugmayeriae*, a parasitoid (of the nymph) not previously recorded from Norfolk. Full details of the discovery can be found in Money (2023).

Acknowledgements

We are grateful to the following:

The Darwin Tree of Life project (via the Enabling Connections fund) together with support from the Earlham Institute's 'Barcoding the Broads' initiative for providing the necessary funds for the initial purchase of equipment and reagents, together with sequencing at Genewiz.

British Mycological Society (DNA Barcoding Project) which is providing funding for sequencing at Aberystwyth University.

Sam Rowe (Earlham Institute) and Brian Douglas (Bento Bioworks, formerly RBG Kew Mycology team) for running various workshops and providing endless practical advice.

Will Fitch and The Ted Ellis Trust for the extensive use of their workshop facilities.

References

Collins. M. 2022. A success for DNA barcoding. *The Norfolk Natterjack* Aug 2022 158:20.

Harries, D.J., Hodges & J.E.Theobald, T. 2018. A study of the distribution of *Microglossum* species in Wales. *Natural Resources Wales Evidence Report No 255*.

Kibby, G. & Tortelli, M. 2021. *The genus Cortinarius in Britain*. Privately published.

Liimatainen, K., Niskanen, T., Dima, B., Ammirati, J.F., Kirk, P.M. & Kytövuori, I. 2020. Mission impossible completed: unlocking the nomenclature of the largest and most complicated subgenus of *Cortinarius, Telamonia. Fungal Diversity* 104:291–331.

Money, T. 2023. The hidden lives of big-headed flies revealed by DNA barcoding. *The Norfolk Natterjack* Feb 2023 160:6-7.

van Vooren, N., Carbone, M., Sammut, C. & Grupe, A.C. 2019. Preliminary notes on the genus Tarzetta (Pezizales) with typifications of some species and description of six new species. *Ascomycete. org,*11(6):309–334.

Wölfel, G. 1994. Seltene Pilze aus der Gattung Entoloma (Rötlinge) Neubeschreibungen aus Franken. *Natur und Mensch - Jahresmitteilungen der Naturhistorische Gesellschaft Nürnberg e.V.* 93-100.