Heather Windram, Terence Charlston and Christopher Howe

A Phylogenetic Analysis of Orlando Gibbons' Prelude in G

Introduction

Parthenia, published in 1612–13, declared on its title page that it was 'the first musicke that ever was printed for the Virginalls', and the final piece in this dedicatory volume is the Prelude in G by Orlando Gibbons (1583–1625).[1] While the printing of text and vocal music had been facilitated by the development of the movable-type printing press in the 15th century, the notational complexities of keyboard music were ill-suited to this method of printing and, when *Parthenia* was printed from engraved metal plates,[2] it was the first book of keyboard music to be published in England.[3] Printed keyboard material became more readily available only with the publication of popular collections: notably those of John Playford and John Carr, later in the century. The complexities involved in the printing of music resulted in the continued propagation of keyboard music by handwritten manuscript throughout the 17th and into the 18th centuries. Even though much manuscript music may have been lost or even destroyed over the intervening years, there is still a considerable extant body of manuscript source material which preserves pieces of music from this period, often in multiple manuscript copies, known collectively as a 'tradition'.

During the process of copying a musical text, deliberate or unintentional changes (or variants) were almost inevitably introduced by the copyist and these changes often resulted in audibly different versions of the same work. Introduced variants were liable to be propagated in subsequent cycles of copying, as observed by Thomas Morley in his Plain and Easy Introduction to Practical Music (1597):[4]

> *...for copies passing from hand to hand a smal oversight committed by the first writer, by the second will bee made worse, which will give occasion to the third to alter much both in the wordes and notes, according as shall seeme best to his owne judgement, though (God knowes) it will be far enough from the meaning of the author, so that errors passing from hand to hand in written copies be easilie augmented,....*

An analysis of the pattern of these inherited variants can give an insight into the transmission or copying history of the tradition. The transmission of music is inherently complex, as it involves not only the written text, but also a performance element, and the performance is as much part of the cultural artefact as is the musical text itself. While the text may be passed down in a series of written copies, the performance remains ephemeral. However, although not preserved in itself, the performance element does not necessarily vanish without trace. The surviving musical texts form a record of the transmission pathway and a comparison between the extant texts may indicate the development of performance interpretation over a period of time. Consideration of variants and transmission history forms a vital part of modern editorial practice and guides historically informed musical performance.

It is possible to use the distribution of variants in a manuscript or early-print tradition to determine the relationships between the copies (termed sources or witnesses) and to ascertain on the basis of shared variants which sources were copied from the same earlier version (exemplar) and thereby to deduce a 'family tree' or stemma. This approach to textual analysis has been taken for many years by scholars of literary texts, and less frequently by scholars of some musical texts. It is called *stemmatics* and to a large extent derives from the pioneering work of the 19th-century German scholar Karl Lachmann,[5] especially as implemented and detailed by Paul Maas in his book *Textkritik*.[6]

It may seem at first glance that there is very little in common between the copying of a text and the processes involved in biological evolution, and yet there are important similarities of concept. The evolution and divergence of species depends on the accumulation of errors, or mutations, in DNA, which forms the genetic code of living organisms. DNA is built up of a chain, or sequence, of units called nucleotides (the four types of nucleotide are identified by the letters A, T, C and G) and, as cells divide, the sequence is copied. Errors (mutations) can happen during copying, producing a daughter molecule with a slightly different sequence. When the daughter molecule is copied, the errors are propagated. As species diverge during evolution, their DNA sequences become increasingly different (diagram 1). This generation and propagation of mutations in DNA molecules during replication has clear parallels with the occurrence and transmission of textual variants, as both systems involve the incorporation of heritable changes during the copying process.

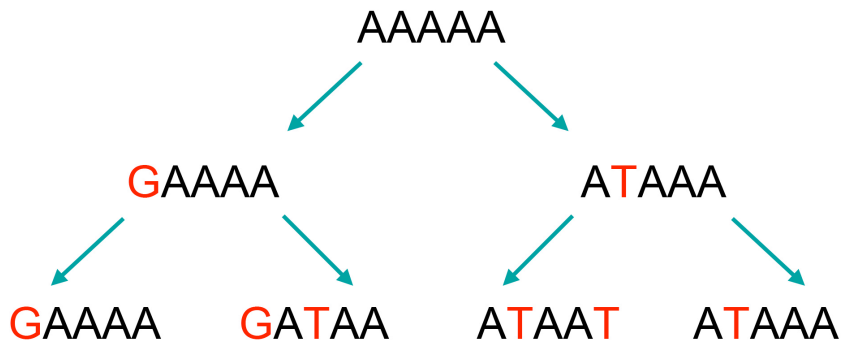


Diagram 1 As species diverge in evolution, differences accumulate between their DNA sequences. The diagram shows a hypothetical example in which divergence of an ancestral species into two is associated with two mutations. Each species then diverges into two again, with an additional mutation occurring in one of each pair of lineages.

Biologists and textual scholars face a common problem of how to understand the relationship between either organisms or texts from the distribution of these inherited errors. With a small amount of text, in a very few witnesses, it might be possible to analyse all variation manually. However, as the amount of text and number of sources to be considered increases, the difficulty of the transmission analysis increases hugely.[7] Improvements in methods for DNA sequencing have resulted in the availability of large quantities of sequence data for analysis, and evolutionary biologists have developed a range of powerful *phylogenetic* computer programs to infer the relationships among different species from the number and pattern of mutations in the DNA. The DNA sequence information is supplied to the computer program as a table (matrix) with each row of characters representing the DNA sequence in one of the species. Each column represents an individual location in the DNA sequence, and specifies what is present at that location for each species under consideration. The phylogenetic algorithms use the distribution of changes in the DNA sequences to indicate the species' evolutionary relationships (which species share common ancestors to the exclusion of other species) and to produce a branching evolutionary tree showing these relationships. Some species may have multiple affiliations as a result of hybridisation of the DNA from more than one parental species, and certain programs are able to depict these multiple affiliations between species as a network rather than as a branching tree.

The principle of copying with incorporation of changes underlies both phylogenetic and stemmatic analyses and, over the last 15 years or so, several projects have applied

computer programs from phylogenetic analysis to studies of the transmission histories of textual traditions. Collaborations with literary scholars have resulted in phylogenetic analyses of a range of textual traditions, such as Chaucer's *Canterbury Tales*,[8] the German legend *Parzival*,[9] Dante's political philosophy treatise *Monarchia*,[10] manuscripts of the Finnish legend of St. Henry[11] and the poetry of the English 17th-century poet, Robert Herrick.[12] Analyses of artificially prepared 'traditions' of known copying history have indicated both the accuracy and some limitations of the approach[13] and have driven the development of new algorithms.[14] The methodology has also been shown to be applicable outside textual criticism and evolutionary biology,[15] having been very successfully applied to studies in the evolution of languages,[16] oral folk traditions[17] and cultural artefacts (e.g. the transmission of the patterns of Turkmen carpets).[18]

Musicologists have also long studied the variation introduced during transmission of musical works.[19] Stemmatic analysis has been used, for example, in studies of the Aquitanian Versaria[20] and the organ music of Christopher Gibbons (the second son of Orlando),[21] and the significance of serial recomposition in variant transmission has been discussed by Herissone.[22] Although phylogenetic methods have been shown to be valuable tools in the analysis of the transmission of textual sources, very little work has been done on applying these computer-based methods to musical text,[23] although such an approach has been used for studies of rhythm,[24] and cluster analysis has been applied to some manuscripts from the *Well-Tempered Clavier II* tradition,[25] and used to study the relationship between folk music melodies and population genetics.[26] In this paper we explore the feasibility of applying the methods of phylogenetic analysis to a music tradition, with a preliminary analysis of the Prelude in G by Orlando Gibbons. This piece is a typical scalic prelude - a genre popular amongst 17th-century English keyboard players. Its preservation in multiple manuscript and early-print witnesses makes it an ideal test case for phylogenetic analysis.

Gibbons' Prelude in G

1 The bust of Orlando Gibbons by Nicholas Stone (1626) in Canterbury Cathedral (Conway Library, The Courtauld Institute of Art, London)

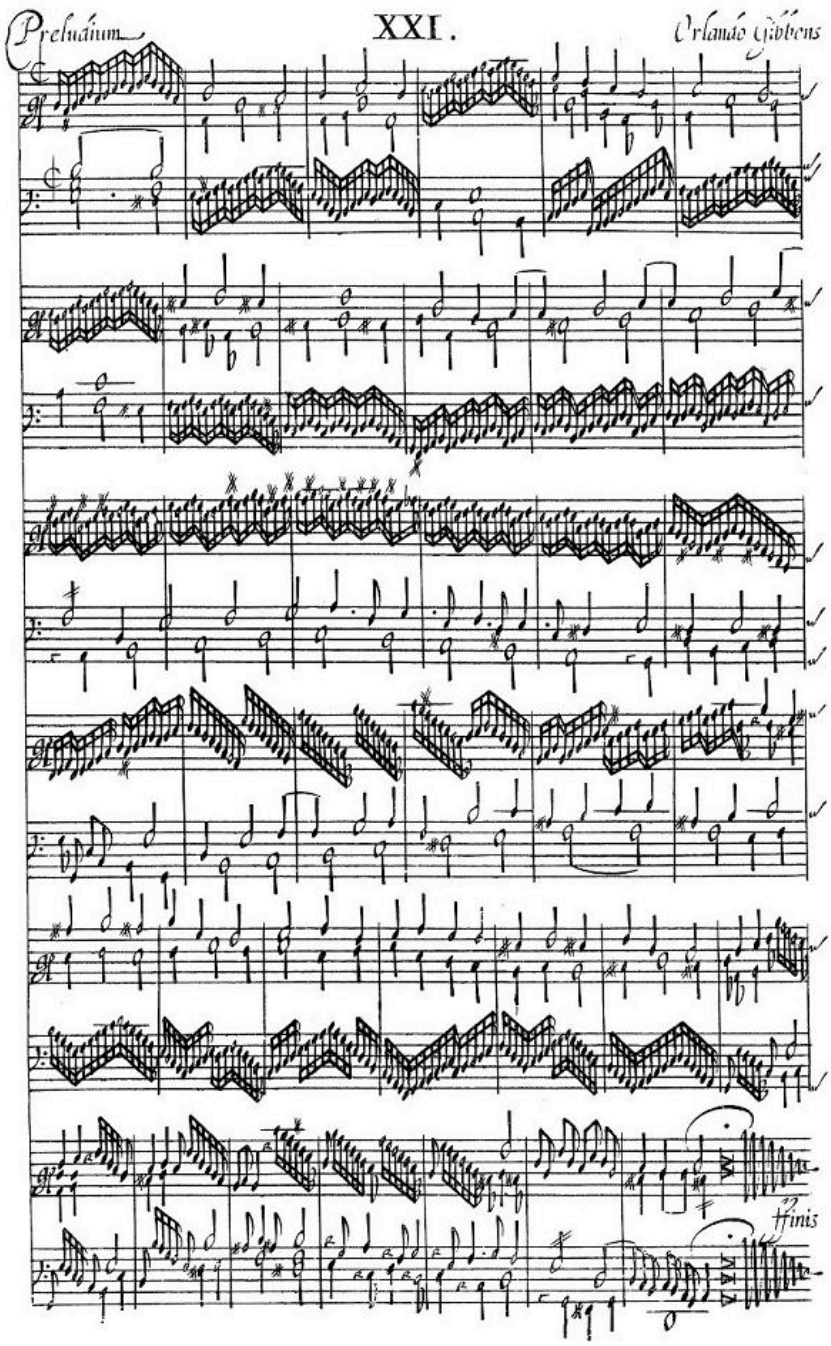
The high esteem in which Orlando Gibbons (illus.1)[27] was held as a keyboard player is indicated not only by contemporary accounts,[28] but also by the inclusion of his keyboard compositions in *Parthenia*.[29] The volume was compiled to celebrate the betrothal of Princess Elizabeth, daughter of James I, and Prince Frederick, elector Palatine, who were married in 1613. Gibbons was in his late twenties at the time of its compilation and he was by far the youngest of the three contributors. His pieces (nos. 16–21) are preceded by 15 compositions by his older contemporaries, William Byrd (1–8) and John Bull (9–15). Gibbons' Prelude in G, the last item in the volume, proved to be an exceptionally popular piece[30] and it survives in 16 extant sources from the 17th and early-18th centuries (see Table 1). Other than the *Parthenia* print, the sources are all manuscript copies – one of which, Cfm Mu Ms 653, bears the added comment:[31]

"*This is N^{o} 21 in the Parthenia & was a favorite Lesson(s) for upwards of a Hundred years*"

In *Parthenia* the piece consists of 38 bars of music (illus.2),[32] with a final chord not separated by a bar-line from the previous bar. Running semiquaver passages move from hand to hand, and are accompanied in the other hand by a two- or three-part texture in longer note values and formed from chords and sequential diatonic counterpoint.

Table 1 The extant sources of the Prelude in G by Orlando Gibbons

Source	Page/Folio/ Number	Abbreviation in paper	Dates[33]
Parthenia – Engraved by William Hole, for Dorothy Evans, and sold by G. Lowe. London.	Piece No. 21	*Parthenia*	1612–1613
GB-London, Royal College of Music, MS 2093	ff. 9v–11v	Lcm 2093	1660s–1670s
GB-Oxford, Christ Church Library, Mus. 47	pp. 43–45	Och Mus. 47	1670s
GB-Oxford, Christ Church Library, Mus. 89	pp. 304–306	Och Mus. 89	1620s
GB-London, British Library, Add. MS 31403	ff. 4v–5r	Lbl Add. 31403	1630s
US-New York, Public Library, Drexel MS 5612	pp. 102–103	NYp Drexel 5612	1620–1660
GB-Cambridge, Fitzwilliam Museum, Mu. Ms. 653	pp. 82–83	Cfm Mu Ms 653	1712–1715
GB-London, British Library, Add. MS 22099	ff. 16v–17r	Lbl Add. 22099	c.1705
D-Berlin, Staatsbibliothek zu Berlin - Preußischer Kulturbesitz, Ms. Lynar A 2	pp. 45–46	B Ms. Ly A2	1620s
GB-London, British Library, Add. MS 23623	ff. 19v–21r	Lbl Add. 23623(i)	1628
GB-London, British Library, Add. MS 23623	ff. 104r–105v	Lbl Add. 23623(ii)	As above
F-Paris, Conservatoire National de Musique (in Bibliothèque Nationale), Rés 1186 bis I	pp. 5–9	Pc Rés 1186 bis I	c.1680
GB-London, British Library, MS Mus. 1	ff. 4r–5r (40v–39v vol. reversed)	Lbl MS Mus. 1	1690s
J-Tokyo, Tokyo College of Music, Ohki Collection, Nanki Music Library, MS N-3/35	ff. 21v–22r	Tn MS N-3/35	1707–1720s
GB-Haslemere, Carl Dolmetsch Library, HAdolmetsch II e. 17	pp. 50–48 (vol. reversed)	HAdolmetsch II e. 17	1680s
US-New Haven, Yale University, Irving S. Gilmore Music Library, Filmer MS 17	ff. 6v and 5v	NH Filmer MS 17	Early 18thC



2 Orlando Gibbons' Prelude in G from *Parthenia* (Image of 1613 edition reproduced from Performers' Facsimiles, 1, piece XXI; by arrangement with Broude Brothers Limited / Performers' Editions)

Data preparation

The phylogenetic computer algorithms cannot work directly with the musical text, but require the variant information to be 'encoded'; in other words, to be converted into a matrix, containing all of the variant information present in all of the sources. There is no distinction made between manuscript and early-print sources, which are given manuscript status and treated identically, with neither category being given any priority or weighting. For this trial analysis, the encoding of the musical data involved a two-stage process.

Firstly, the variations present in all of the sources were identified and categorised. At each variant location, the reading present in each source was recorded and classified into the appropriate category, for example, 'pitch', 'note pattern' (e.g. the choice of perhaps a semibreve or two tied minims, stem direction or beaming), 'ties' or 'ornamentation'. However, the written music serves as a guide or indicator for performance. Some variation may be silent, for example the semibreve and the two tied minims would be performed identically, but omission of the tie in another source would result in an audible performance variation. So, an additional category of 'rhythm' was used to record variation in the audible note entrances - in this example, the semibreve and tied minims would be recorded as identical while the two separate minims would be recorded as a difference, as a new sound would enter at the third crotchet beat. Additionally, all variants were also recorded as right hand (upper stave), left hand (lower stave) or both (affecting both staves e.g. key signature). Fingering was not included in the dataset, as it may have been added at any time to the texts, and so may not reflect the copying history of the text.

Secondly, these variant data were recorded as a matrix (termed a Nexus file), as shown in diagram 2, with each row representing the variant information from one source, and each of the 610 columns representing a variant location in the musical text.

Source	Columns 1–34 of Nexus file
Parthenia	00000000000000000000000000000000
Lcm 2093	11001010000000000000000000000000
Och Mus. 47	00002110110011111010111?1010100010
Och Mus. 89	00003110200100000000000000200000
Lbl Add. 31403	220010?10000200000010000010120000
NYp Drexel 5612	33000020110000000000000000000000
Cfm Mu Ms 653	22110011110011110210121111?031?010
Lbl Add. 22099	40100011?1111110210121111110410110
B Ms. Ly A2	00000021110000000000000000000000
Lbl Add. 23623(i)	540001?00010112110000000000000001
Lbl Add. 23623(ii)	640001?01110111100000000000000000
Pc Rés 1186 bis I	20000011000011211300030000000200000
Lbl MS Mus. 1	21000010000111110101400101050001O
Tn MS N-3/35	75010011110011110410151111110310010
HAdolmetsch II e. 17	81100010000111110000000210610000
NH Filmer MS 17	21100010000020000000000000200000

Diagram 2 Section of Nexus file for the 16 sources of Gibbons' Prelude in G. Each row represents the data from one source and each column represents the readings at a site of variation in one or more sources. Column 19 is highlighted in red (see text).

This information simply serves to align the sources according to their readings at each variant location. The choice of symbols used in the Nexus file is not significant. In this analysis we chose to score the reading in the *Parthenia* text as '0', and all sources sharing the same reading at that location would also score as '0'. The first variant reading in that column would score as '1' and the next would score as '2'. This can be illustrated by e.g. column 19 (in red in diagram 2) which represents a 'rhythm' variant in the S voice of bar 2, where a '0' represents sources, including *Parthenia*, which have a single sound lasting for two crotchet beats, while a '1' represents sources that have a new sound entering on the second crotchet beat. The symbols do not indicate numerical values or weighting of the data, and the variants could alternatively have been scored for example as 'a', 'b' and 'c' etc. In locations where

there is missing information e.g. the pitch of a note which is not present in one of the sources, then a '?' is recorded. The '?' symbol is not included in the calculations performed by any of the phylogenetic programs. The dataset also contains lists of the column numbers (or characters) indicating which columns represent each category of variant (e.g. 'rhythm' or 'pitch'), and these 'character sets' can be used to include or exclude categories of variants from the analyses. The Nexus file can be used directly by the phylogenetic computer programs.

Phylogenetic analyses

A wide range of phylogenetic algorithms has been developed for analysis of biological datasets. However, for this work we restricted our analyses to two of the methods which have also been successfully used to analyse a range of literary traditions.

Maximum parsimony[34] (MP) creates a phylogenetic tree according to the principle that it is more 'parsimonious' for a variation/mutation that is found in multiple sources/species to have occurred only once, and to have been inherited from a common ancestor, rather than for it to have been introduced independently multiple times in different sources/species. The algorithm therefore finds the tree that requires the minimum number of mutation events. The species/sources are grouped according to their shared derived characters, and the resultant tree shows 'maximum parsimony' in comparison with all other possible trees drawn from that set of sources. A statistical technique known as 'bootstrapping' can be used, to give a measure of the confidence in assigning a set of sources/species to a particular group to the exclusion of others. This is typically expressed as a percentage value indicated on the branch that unites those sources/species.[35]

Neighbor-Net[36] (NNet) uses the dataset to create a distance matrix of all possible pairs of sources/species, with the distance being calculated from the number of sites at which the two sources differ. This distance matrix is then used to construct a tree or network. Pairs of species/sources are selected iteratively, and pairs of pairs with one common node are agglomerated, allowing a network of relationships to be constructed. The resultant network can represent multiple affiliations between sources.

Both of these methods provide phylogenetic trees that are unrooted (i.e. the oldest point on the tree is not inferred in the absence of any other information or assumptions) and therefore do not indicate the origin of the tradition. Although this could be one of the extant witnesses, it could alternatively be a lost source and therefore not part of the analysis.

Results

The MP analysis (diagram 3) of the Prelude dataset shows that the sources divide into several well-defined groupings with good statistical support. One branch of the tree (group A) includes the *Parthenia* version of the piece, which groups most closely with B MS Ly A2 and then with NYp Drexel 5612. This group then extends to include the Lcm 2093 copy. A second group (B) consists of Och Mus. 89 and the two copies found in Lbl Add. 23623. The two texts in Lbl Add 23623, although clearly not identical, are sufficiently similar to group together with 100% bootstrap support, and Och Mus. 89 is often linked in the use of accidentals with the Lbl Add. 23623 texts. Only the first 30 bars of Och Mus. 89 are included in the analysis as it preserves an alternative ending, based on doubled note values. An ending with doubled note values is also found in Lbl Add. 31403.[37] The endings for these two sources are similar, but not identical and are not included in this overview analysis. The first 30 bars of Lbl Add. 31403 groups strongly (100% bootstrap support) with Pc Rés 1186 bis I (Group C) and the sources share a number of exclusive variants, often involving the 'note pattern' category of variants. A final grouping of seven sources (Group D) is supported at 99%.

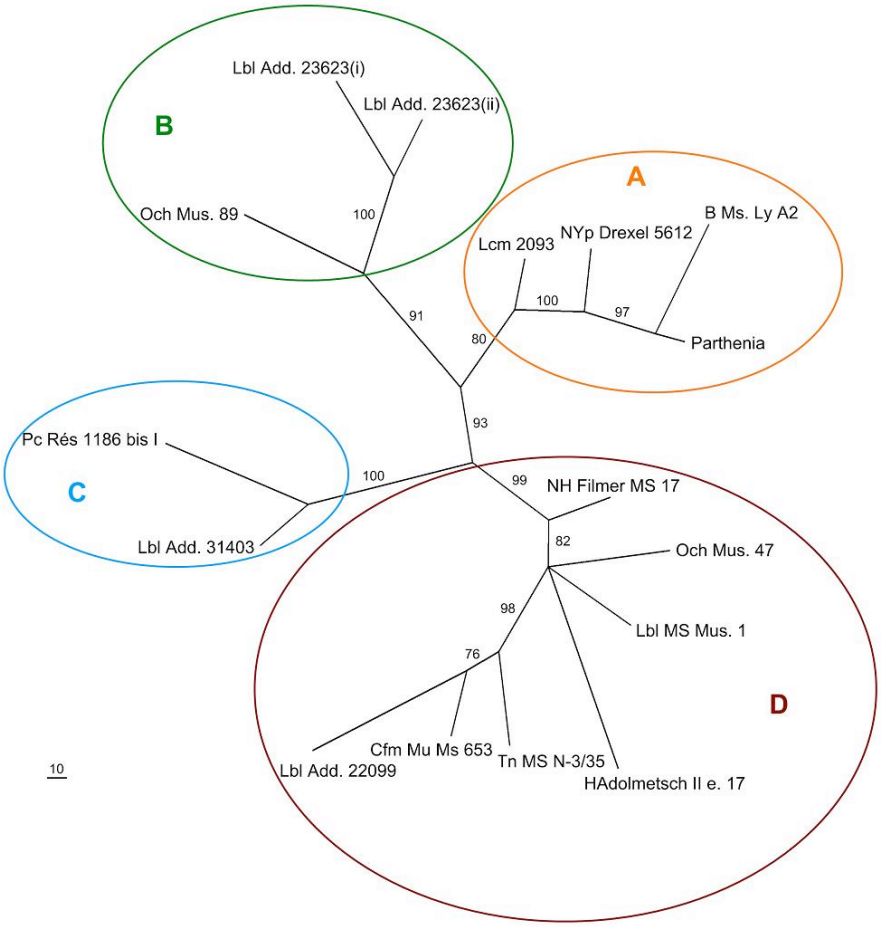


Diagram 3 Radial-tree presentation of the MP analysis of the Prelude dataset. The numbers on the branches indicate the bootstrap support for each branch. The branch lengths are proportional to the number of substitutions inferred (as shown in the scale bar). The rings superimposed on the tree indicate phylogenetic groupings.

The NNet tree (diagram 4), although constructed using a fundamentally different approach, agrees strongly with the MP analysis. The *Parthenia* group of sources is again apparent, with Lcm 2093 appearing more loosely linked in this analysis. As the NNet tree allows multiple affiliations to be visualised through regions of networking, it is possible to see that there is some linkage between Och Mus. 89 and Lbl Add. 31403, indicating a possible relationship between these texts that cannot be fully represented in MP which can only show a single affiliation for each source, and suggesting a link between the B and C groups of sources The sharpest division in the NNet tree is seen in the separation of the group of seven sources that form Group D.

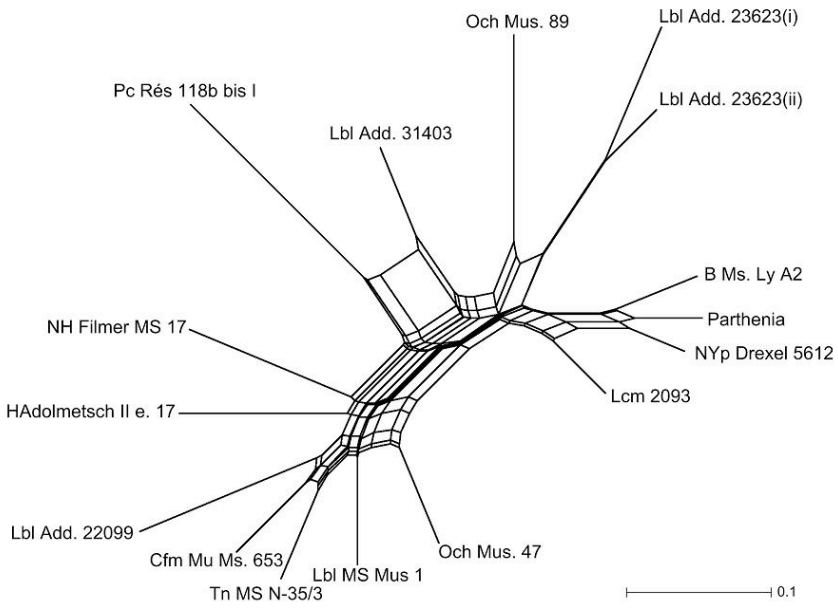
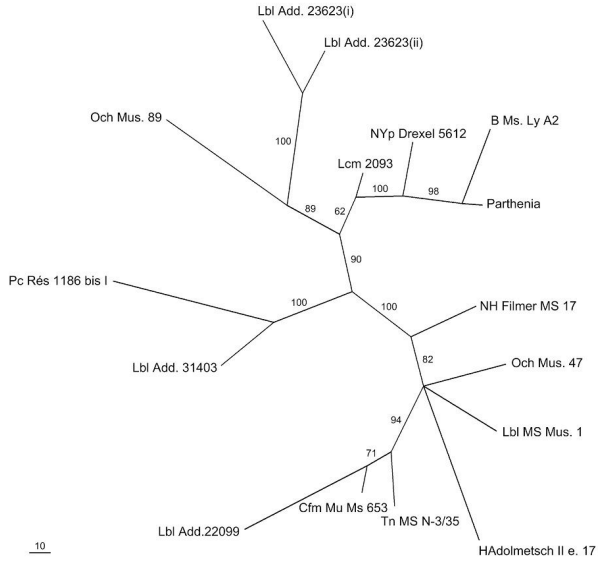


Diagram 4 NNet analysis of the Prelude dataset. The edge lengths are proportional to the number of differences between sources (expressed in the scale bar as the mean character difference).

In the Prelude Nexus file, there are no data beyond bar 30 (column 430 out of 610) for the sources Och Mus. 89 and Lbl Add. 31403, as the endings with doubled note values cannot be compared directly with the endings in the other manuscripts.[38] It is possible that this discrepancy may affect the affiliations and groupings of the texts. One of the valuable features of the phylogenetic analyses is that it is possible to re-run analyses speedily, including only chosen sections of data, or selected sources, or specific types of variants (e.g. pitch or rhythm variants, in the case of music data). We repeated the analysis using only columns 1–430, to compare the data for the region that is common to all of the sources.[39]

a)



b)

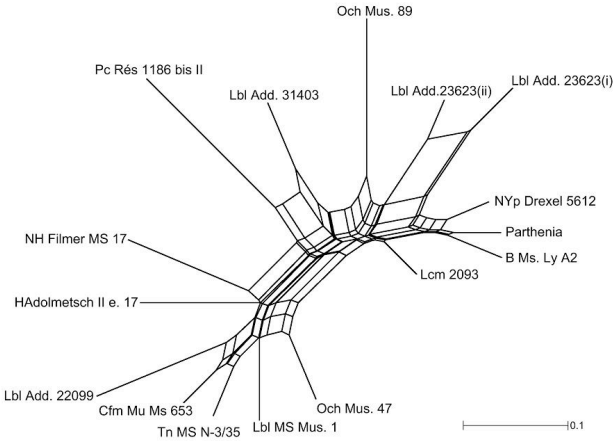


Diagram 5 a) MP and b) NNet analyses of columns 1–430 in the Prelude dataset.

The MP tree of 1–430 (diagram 5a) shows the same groupings as seen in the analysis of the full dataset. There is slightly less bootstrap support for the association of Lcm 2093 (62%), but the grouping still forms part of this consensus tree. All other groupings are supported with almost unchanged bootstrap percentages. The NNet analysis (diagram 5b) is also very similar to that of the full dataset, although Och Mus. 89 is positioned more centrally between Lbl Add. 31403 and the two Lbl Add. 23623 sources, with clear linkages to both groups. The clearest split is again between the same group of seven sources and the rest of the tradition. This group is now supported at 100% in the MP tree.

The running semiquavers passages shift from hand to hand throughout the piece, with the other hand always providing a chord-based accompaniment. By analysing character sets selected from within the 1–430 dataset, it is possible to perform the analyses on only the variants from either the 'semiquaver' hand (156 col.) or the 'accompaniment' hand (269 col.). A comparison of the MP trees is shown in diagram 6. The MP tree shows that the groupings seen in the analyses of the full dataset and of columns 1–430 are largely retained in the analysis of the accompanying hand (although with Och Mus. 89 showing its association with Lbl Add. 31403), with the D group of seven sources being supported at 100% (diagram 6a). However, the structure and grouping is mostly lost in the analysis of the semiquaver hand, which results in a tree which contains generally low support values, and with many sources forming unresolved 'stars' in which the bootstrap support for any structure is less than the threshold of 50% (diagram 6b). The variants from the runs of semiquavers include many stem direction and beaming differences, and only the diligent scribes of NYp Drexel 5612 and B Ms. Ly A2, copying directly from *Parthenia*, have faithfully copied these details such that a well-supported grouping is still visible in the analysis. These results indicate that the phylogenetic signals responsible for the groupings seen in the full dataset and 1–430 dataset are largely contained in the variants from the chord-based accompaniment. The variation in the semiquaver melody is more random and does not indicate a consistent pattern of relationship between the sources. These results are fully consistent with Herissone,[40] who points out that the variation between the Prelude sources occurs mostly in the 'non-active' accompanying hand.

a) b)

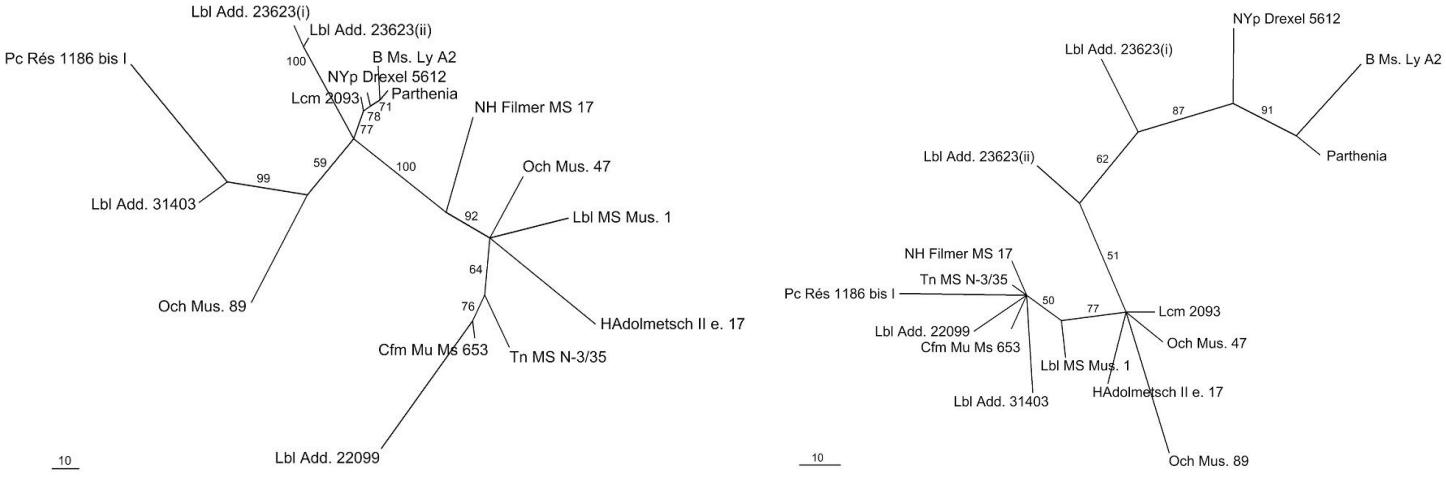


Diagram 6 MP analyses of a) the accompanying-hand variants and b) the semiquaver-hand variants.

In addition to a large number of locations where variants occur in only one (or a few) of the sources, there are five bars within the first 30 bars (represented by columns 1–430) in

which the readings of the accompanying hand can be divided into two main versions which are found across the tradition.[41] The sources sharing a particular version of a bar are not identical, but they share the key features of that version. An example showing the variation in one of these bars (bar 20 left hand) is given in diagram 7. For each of these five bars, Table 2 gives a summary of the version found in each of the sources, with the *Parthenia* version numbered 1.

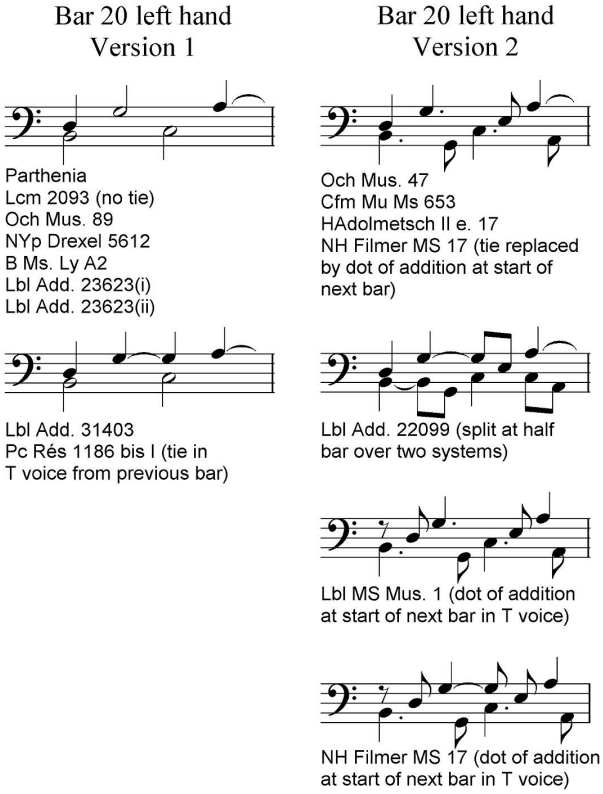


Diagram 7 The variant readings in bar 20 left hand. The reading in each source falls into one of two main versions, although the readings within each version are not all identical.

Table 2 The distribution of the variant versions in the sources of the Prelude in G

Source	**Version in bar/hand:**				
	2/rh	**9/rh**	**20/lh**	**21/lh**	**22/lh**
Parthenia	1	1	1	1	1
Lcm 2093	1	1	1	1	1
Och Mus. 47	2	2	2	2	2
Och Mus. 89	1	1	1	1	1
Lbl Add. 31403	1	1	1	1	1
NYp Drexel 5612	1	1	1	1	1
Cfm Mu Ms 653	2	2	2	2	2
Lbl Add. 22099	2	2	2	2	2
B MS Ly A2	1	1	1	1	1
Lbl Add. 23623 (i)	1	1	1	1	1
Lbl Add. 23623 (ii)	1	1	1	1	1
Pc Rés 1186 bis I	1	1	1	1	1
Lbl MS Mus. 1	2	2	2	2	2
Tn MS N-3/35	2	2	2	2	2
HAdolmetsch II e. 17	1	2	2	2	2
NH Filmer MS 17	1	2	2	2	2

With very few exceptions, the sources consistently have either the *Parthenia* version or the alternative version at each of the five locations. HAdolmetsch II e. 17 and NH Filmer MS 17 are the only two sources to contain a mixture of the versions and they both match the *Parthenia* version at the first location, but otherwise carry the alternative version. From the readings in this table, the sources may be considered to belong to groups '1' or '2'. This grouping of the sources can be superimposed on the NNet of the full dataset as shown in diagram 8.

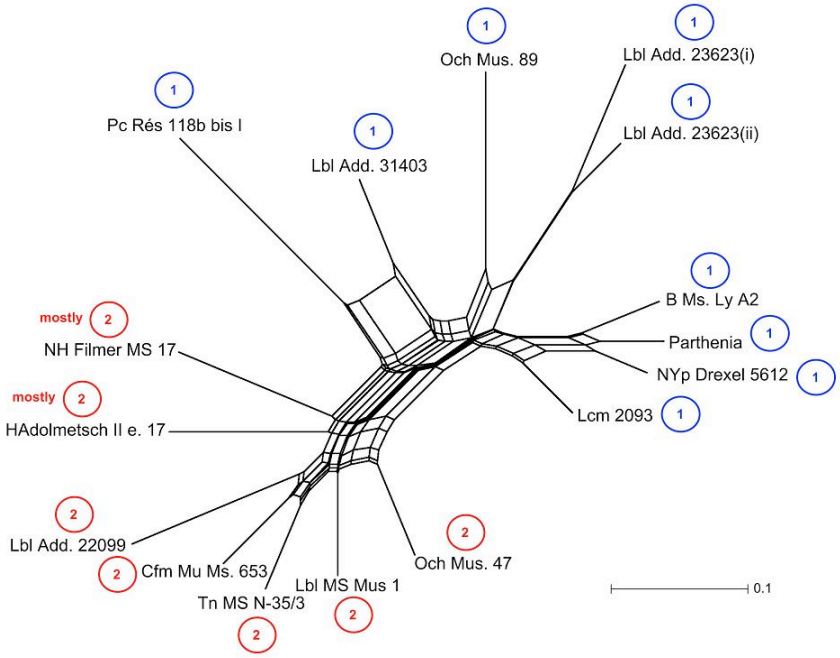


Diagram 8 The variant version grouping superimposed on the NNet analysis of the full dataset.

The grouping of sources according to their variant versions at these locations matches exactly with the two main groupings identified in the NNet analysis. Group 1 corresponds with the groupings already identified as A, B and C, while group 2 consists of the seven sources which formed group D.

Discussion

Phylogenetic analysis of the Prelude in G was performed on the variant data from all extant sources of the piece including the 1613 print and manuscripts dating from the early-17th century to the mid-18th century. The results of the phylogenetic analysis indicate several features about the relationship between the sources.

The sources B MS Ly A2, NYp Drexel 5612 and Lcm 2093 tend to cluster around the *Parthenia* print, with Lcm 2093 consistently at a slightly greater distance from *Parthenia* than the other two sources, as seen in the NNet analyses. Within this group there is no particular substructure, and it would seem that these sources may have been copied directly and independently from the original *Parthenia* publication. Manuscript Och Mus. 89 shows some relationship with both the Lbl Add. 23623 sources and Lbl Add. 31403, which is the only other source to present an ending to the piece in doubled note values. Other than this common approach to the ending, there is no overwhelming association between the two sources, although the linkage that is apparent is still visible when only the first 30 bars of all sources

are compared. The analysis raises questions about the relationship between these texts and indicates that further examination would be fruitful.

The phylogenetic analyses also indicate that the extant sources can be divided into two main groups (1 and 2 in diagram 8), which is fully supported by a comparison of the variant versions present in a number of bars in the accompanying 'inactive' hand. In each case, the differences between the two versions are fairly trivial; however, it is of interest that each source contains, almost exclusively, either version 1 or version 2 readings throughout. Only HAdolmetsch II e. 17 and NH Filmer MS 17 (an incomplete copy split between two non-sequential pages) have one variant bar based on the *Parthenia* version with the other variant bars belonging to the second grouping. These results indicate that there may have been two main versions of the Prelude text circulating, although they do not preclude the possibility that the two versions may both have originally derived from the *Parthenia* text. The older extant sources (early- to mid-17th century) are included within group 1, while group 2 contains mostly sources dating from the latter part of the 17th century and into the first half of the 18th century. Version 1 is seen at the latest in the 1680s copy in Pc Rés 1186 bis I, while the later version is first seen in Och Mus. 47 and then is found in all of the sources from 1690 onwards.[42] The development of the second group of texts may reflect changing performance preferences in the century after the original publication of the Prelude in *Parthenia*.

For a small tradition, it is clearly easier to undertake a manual analysis, where a full range of variants can be tabulated and recalled, and where metadata such as paper type, bindings and dates of sources can be considered in the analysis. As a tradition increases in size, it becomes increasingly hard for all variants and all possible inter-relationships to be considered. One benefit of the computer-based methods is that they are able to give a numerical overview of the relationships as indicated by the sum total of variation in the dataset. There is no need to focus on a few favoured variants, the selection of which can result in a circular argument concerning authorial intention,[43] although any categories of variants considered to be non-substantive for a particular tradition can be excluded from the analysis if this is desired. There is no automatic weighting of data, but weightings of variant types can be included if this should be required. A large quantity of data can be assimilated and analysed at the same time, and it is possible to obtain a full overview of the signals in the data in a way that would not be possible manually. Moreover, the phylogenetic analysis can be used to probe the tradition and raise pertinent questions e.g. the possibility that there are two lines of transmission in the Prelude. One great asset of the phylogenetic approach is that, having prepared the basic dataset, it is possible to analyse and re-analyse the data very quickly, possibly including only certain categories of variants, selected sources, or specific subsections of the data, and to obtain the results from the analyses within the few minutes required for the running of the phylogenetic algorithms.

However, it must be stressed that we advocate a sensitive use of phylogenetic methods, in conjunction with a more traditional approach. No computer analysis can replace the expertise of the musicologist or textual scholar, who considers a mass of background information in addition to the purely numerical patterns of variation and, where the two approaches conflict, it is essential to take into account the detailed knowledge of the musicological or textual scholar.

The analysis of the Prelude in G by Orlando Gibbons forms a test case for the application of the methods of phylogenetic analysis to a study of the transmission history of a music manuscript and early-print tradition. The phylogenetic methods give statistically robust conclusions, suggesting that there is genuine historical signal in the data. It is also significant that the different phylogenetic methods (MP and NNet) give consistent results, which is taken as an indicator of reliability in phylogenetic analysis with biological sequence data. Previous collaborative projects involving literary textual traditions have indicated that the real strength

of the phylogenetic analysis is its value as an additional tool to add to the toolbox of the textual scholar studying the transmission history of suitable literary traditions, and this study indicates that it should be feasible to extend this approach to the study of suitable music traditions.

Acknowledgements

We are grateful to Andrew Woolley for his helpful comments on a draft of this paper.
We thank the Royal College of Music for allowing access to GB-London, Royal College of Music, MS 2093. The manuscript GB-Haslemere, Carl Dolmetsch Library, HAdolmetsch II e. 17 was accessed by kind permission of Jeanne Dolmetsch.

---

[1] William Byrd, John Bull, Orlando Gibbons, *Parthenia or the maydenhead of the first musicke that ever was printed for the virginalls*.... (London, 1612–13).
[2] It has been suggested that the engraving may have been made from Gibbons' own autograph . G. Hendrie, 'The keyboard music of Orlando Gibbons (1583–1625)', *Proceedings of the Royal Musical Association*, lxxxix (1962), pp.1–15, at p.3.
[3] For a description of printing from moveable type, engraving (and lithography) see A. Devriès-Lesure, 'Technological Aspects', *Music publishing in Europe 1600–1900*, ed. R. Rasch (Berlin, 2005), pp.63–88.
[4] Thomas Morley, *A plaine and easie introduction to practicall musicke set downe in forme of a dialogue*..., (London, 1597), p.151.
[5] A history and description of the Lachmannian method is given in S. Timpanaro, *La genesi del metodo del Lachmann* (Florence, 1963; rev. edn. Padua, 1985).
[6] P. Maas, *Textkritik* (Leipzig & Berlin, 1927) translation B. Flower as *Textual criticism* (Oxford, 1958).
[7] With four sources there are three possible unrooted tree topologies, and with five sources there are 15. However, by the time there are ten sources, there are over 2 million possible unrooted trees, rising to over 34 million possible tree topologies if all rooted options are considered.
[8] A. C. Barbrook, C. J. Howe, N. Blake and P. Robinson, 'The phylogeny of *The Canterbury Tales*', *Nature*, cccxciv (1998), p.839.
[9] M. Stolz, 'New philology and new phylogeny: aspects of a critical electronic edition of Wolfram's *Parzival*', *Literary and Linguistic Computing*, xviii (2003), pp.139–50.
[10] H. F. Windram, P. Shaw, P. Robinson and C. J. Howe, 'Dante's *Monarchia* as a test case for the use of phylogenetic methods in stemmatic analysis', *Literary and Linguistic Computing*, xxiii (2008), pp.443–63.
[11] T. Roos, T. Heikkilä and P. Myllymäki, 'A compression-based method for stemmatic analysis', *Proceedings 17th European Conference on Artificial Intelligence*, (2006), pp. 805–6.
[12] H. F. Windram, C. J. Howe and R. Connolly, 'Thinking 'bibliogeographically': phylogenetic analyses and systems of manuscript circulation', *English Manuscript Studies*, xviii (2013), pp.215–51.
[13] M. Spencer, E. A. Davidson, A. C. Barbrook and C. J. Howe, 'Phylogenetics of artificial manuscripts', *Journal of Theoretical Biology*, ccxxvii (2004), pp.503–11; T. Roos, T. Heikkilä, 'Evaluating methods for computer-assisted stemmatology using artificial benchmark data sets', *Literary and Linguistic Computing*, xxiv (2009), pp.417–33; C. Macé, P. Baret, P. Robinson, 'Testing methods on an artificially created textual tradition', *Linguistica Computazionale*, xxiv–xxv (2006), pp.255–83.
[14] T. Roos and Y. Zou, 'Analysis of textual variation by latent tree structures', *Proceedings of 2011 International Conference on Data Mining*, eds. D. Cook, J. Pei, W. Wang, O. Zaïane and X. Wu, (2011), pp.567–76.
[15] For a review of the field see C. J. Howe and H. F. Windram, 'Phylomemetics—Evolutionary analysis beyond the gene', *PLoS Biology*, 9(5) (2011), e1001069. doi:10.1371/journal.pbio.1001069.
[16] R. D. Gray, A. J. Drummond and S. J. Greenhill, 'Language phylogenies reveal expansion pulses and pauses in Pacific settlement', *Science*, cccxxiii (2009), pp.479–83.
[17] J. J. Tehrani, 'The phylogeny of Little Red Riding Hood', *PLoS ONE*, 8(11) (2013), e78871. doi:10.1371/journal.pone.0078871.
[18] J. Tehrani and M. Collard, 'Investigating cultural evolution through biological phylogenetic analysis of Turkmen textiles', *Journal of Anthropological Archaeology*, xxi (2002), pp.443–63.
[19] B. Cooper, 'Problems in the transmission of Blow's organ music', *Music & Letters*, lxxv (1994), pp.522–47.
[20] J. Grier, 'The stemma of the Aquitanian Versaria', *Journal of the American Musicological Society*, xli (1988), pp.250–88; J. Grier, 'Scribal practices on the Aquitanian Versaria of the twelfth century: towards a typology of error and variant', *Journal of the American Musicological Society*, xlv/3 (1992), pp.373–427.
[21] G. Cox, *Organ music in Restoration England: a study of sources, styles and influences* (New York and London, 1989), i, pp.81–90.

[22] R. Herissone, *Musical creativity in Restoration England* (Cambridge, 2013), pp.209–59 discusses serial recomposition, mostly with reference to liturgical music.
[23] A rare example is a study of motets by Thomas Tallis in P. Rapson, *A technique for identifying textual errors and its application to the sources of music by Thomas Tallis* (New York, 1989), but this work pre-dates the development of the sophisticated phylogenetic algorithms now available.
[24] C. Guastavino, F. Gómez, G. Toussaint, F. Marandola and E. Gómez, 'Measuring similarity between Flamenco rhythmic patterns', *Journal of New Music Research*, xxxviii/2 (2009), pp.129–38.
[25] M. Niitsuma, T. Fujinami and Y. Tomita, 'The intersection of computational analysis and music manuscripts: a new model for Bach source studies of the 21st century', *Proceedings of 10th International Society for Music Information Retrieval Conference*, eds. K. Hirata, G. Tzanetakis and K. Yoshii (2009), pp.519–23.
[26] H. Pamjav, Z. Juhász, A. Zalán, E. Németh and B. Damdin, 'A comparative phylogenetic study of genetics and folk music', *Molecular Genetics and Genomics*, cclxxxvii (2012), pp.337–49.
[27] For a biography of Orlando Gibbons see J. Harper 'Orlando Gibbons', *Oxford Dictionary of National Biography*, (Oxford, 2004) online ed, Jan 2008 [http://www.oxforddnb.com/view/article/10598, accessed 4 April 2014].
[28] For example, 'the best finger of that age'. John Hacket, 1592–1670, Bishop of Coventry and Lichfield, in *Scrinia Reserata: a memorial offer'd to the great deservings of John Williams, D.D.* (published posthumously), (London, 1693), Part 1 p.210.
[29] *Parthenia* is also the first English music book issued from engraved plates. There is a copy of the first impression in the British Library (R.M.15.i.15) and at least five other extant copies and/or impressions held in other libraries and collections. D. W. Krummel *English music printing, 1553–1700* (London, 1975), pp.143–5. Hendrie suggests that of all the sources, 'only in *Parthenia* do we sense Gibbons' authority'. G. Hendrie, *Orlando Gibbons: keyboard music*, Musica Britannica XX (London, 1974), Editorial Method p.xv.
[30] C. Bailey, *Seventeenth-century British keyboard sources* (Warren, Michigan, 2003), p.21.
[31] This comment has been added to Cfm Mu Ms 653 by an unknown hand and is followed by the initials J. W (or M). There is no indication of the date of this addition, but it may post-date the ownership of the manuscript by William H. Cummings (1831–1915).
[32] Broude Brothers Ltd. *Parthenia*, Performers' Facsimiles, 1, (New York, 1985).
[33] Some sources were compiled in several sections over a period of time. In these cases, the dates given refer to the section of the source in which the Prelude in G occurs. The following sources were consulted for dating purposes: Bailey, *Seventeenth-Century British Keyboard Sources*, p.58 and p.79; V. Brookes, *British Keyboard Music to c.1660: Sources and Thematic Index*, (Oxford, 1996); British Library, *Search our Catalogue - Archives and Manuscripts*, [http://searcharchives.bl.uk/primo_library/libweb/action/search.do?vid=IAMS_VU2] see entry for Add. 23623 (accessed 7 August 2014); P. Dirksen, 'New perspectives on Lynar A1', *The Keyboard in Baroque Europe*, ed. C. Hogwood (Cambridge, 2003), pp.36–66, at p. 66; J. Milsom, *Christ Church Library Music Catalogue*, [http://library.chch.ox.ac.uk/music/page.php?set=Mus.+89, page last updated 2010, accessed 7 August 2014); A. Woolley, *English Keyboard Sources and their Context, c.1660–1720* (PhD diss., U. of Leeds, 2008), Appendix B, pp.228–76; A. Woolley, 'Manuscript additions to a copy of John Playford's *Select Musicall Ayres and Dialogues* (1652) in the Dolmetsch Library (II. e. 17): a little-known source of late seventeenth-century English keyboard and vocal music', *The Consort: The journal of the Dolmetsch Foundation*, lxvi (Summer, 2010), pp.35–53, at p.40.
[34] Maximum parsimony was performed using the PAUP* software package - see D. L. Swofford, *PAUP*. Phylogenetic analysis using parsimony (*and other methods)* (Sunderland, MA, 2001).
[35] J. Felsenstein, 'Confidence limits on phylogenies: an approach using the bootstrap', *Evolution*, xxxix (1985), pp.783–91. Bootstrapping is not in itself a phylogenetic algorithm, but is a system of sampling with replacement which operates within the software package. This system creates and analyses datasets of identical size to the original dataset, containing columns randomly selected from the original dataset and then replaced in the pool for possible further selection. The data selection and analysis are automatically repeated a large number of times (1,000 in this work), and the trees created by each run are combined to give a consensus tree which shows only groupings that are present in a specified proportion (50% or higher in this work) of the individual trees. A value is assigned to each branch of the consensus tree, representing the percentage of runs in which a particular grouping occurs. Bootstrap values are not given for the terminal branches as, regardless of tree structure, each source is located on its own terminal branch, and so these values would always be 100%.
[36] D. Bryant and V. Moulton, 'Neighbor-Net: an agglomerative method for the construction of phylogenetic networks', *Molecular Biology and Evolution*, xxi/2 (2004), pp.255–65. Neighbor-Net was implemented using the SplitsTree4 software package – see D. H. Huson and D. Bryant, 'Application of phylogenetic networks in evolutionary studies', *Molecular Biology and Evolution*, xxiii/2 (2006), pp.254–67.
[37] Hendrie suggests that the revised ending may be the result of a re-working by Gibbons himself. See Hendrie, *Orlando Gibbons*, Editorial Method p.xv.

[38] In theory such a comparison could be made by halving the note values for the endings with doubled values, however, this would require many subjective assessments and was considered to be outside the scope of this work.
[39] The Prelude in NH Filmer MS 17 is written on two non-consecutive pages and ends after bar 22 (column 298 of 610 in the Nexus file), but variant data for all of the available music is included in the analysis.
[40] See Herissone, *Musical creativity*, pp.370–2.
[41] There is a division into similar groupings in bars at the end of the piece, but these are not considered here as the double note value endings of Och Mus. 89 and Lbl Add. 31403 are not included in this work.
[42] It is always possible for a later source to carry an earlier version of the text, although the converse is not feasible.
[43] Discussed in Windram *et al*., 'Thinking 'bibliogeographically'' at pp. 218–220.