

A phylogenetic analysis of two preludes from J. S. Bach's *Well-Tempered Clavier II*

PHYLOGENETICS developed in biology to study the evolutionary history of, and relationships among, groups of organisms. Computer algorithms have been developed to infer phylogenetic (evolutionary) relationships using biological data, such as DNA sequences, which are too complex to be analysed manually. There are many similarities between inferring evolutionary relationships among groups of organisms using DNA sequences and inferring the copying history of different versions of a written text, and many studies have used phylogenetic methods to infer the transmission history of literary works. The current article tests the application of a phylogenetic method to a large dataset of variant readings for a musical text—namely J. S. Bach's *Well-Tempered Clavier II*, which is well known for the complexity of its source situation. In his 1995 study of all the variant readings found in extant sources of the work, Yo Tomita highlighted the need to 'calculate the complex array of data and measure the degree of uncertainties'.¹ Here, variant data from two contrasting pieces (the Preludes in C major and A major, BWV 870 and 888) are analysed to determine if phylogenetic analysis allows musically and historically plausible insights into their transmission.

Bach's *Well-Tempered Clavier II*

Twenty years after completing the first part of the *Well-Tempered Clavier*, Bach added the final pieces to a second collection of compositions, also comprising 24 preludes and fugues in every major and minor key, and now generally referred to as the second part of the *Well-Tempered Clavier*. However, while the two compilations may have shared similarities of concept and intention, the physical fates

of the two parts are very different. The first part of the *Well-Tempered Clavier* (WTC I) survives as a complete fair copy, bound and dated 1722 on its title-page.² While some of the pieces received minor revisions later in the composer's life, there is no reason to doubt that Bach's fair copy of WTC I broadly records his intended musical text of each piece as it stood in the early 1720s. The title-page carries an elaborate text describing the contents and the didactic purposes for which Bach had made this compilation.³ For the second part of the *Well-Tempered Clavier* (WTC II), no final fair copy survives. Indeed, although the compilation is believed to have been completed by 1742, the earliest surviving complete copy (1744) is found not in Bach's hand, but in that of Johann Christoph Altnickol (1719–59), Bach's student and future son-in-law. Here, for the first time, the collection acquired a title-page modelled on that of the first part:⁴

Second Part of The Well-Tempered Clavier, consisting of Preludes and Fugues in all tones and semitones prepared by Johann Sebastian Bach, Royal Polish and Electoral Saxon Court Composer, Capellmeister and *Directore Chori Musici* in Leipzig.

This inscription offers, at one remove, the first indication that Bach considered the collection as a second part of the *Well-Tempered Clavier*, although it omits any mention of the perceived purpose for which the new collection was compiled. The lack of a complete fair copy in the composer's hand and the consequent implication that WTC II is unfinished has led to a 'consensus among commentators' that the second part 'is less attractive than its predecessor'.⁵ Nevertheless, in this second part, Bach showed

his mastery of newer styles and fugal techniques, with larger scaled and 'more thoroughly worked out' preludes than those of the first part,⁶ and certain self-imposed restrictions such as fugues only in three or four voices.⁷

The sources of WTC II

In addition to the first complete printed editions, published almost simultaneously by Simrock (Bonn, 1801), Nägeli (Zurich, 1801) and Hoffmeister/Kühnel (Vienna and Leipzig, 1802–4), there are well over 100 extant manuscripts preserving some or all of the contents of WTC II. This article gives abbreviated names for the manuscript sources; see [Appendix 1](#) for their full details. Most of the movements in WTC II are included in the main partial autograph manuscript, the so-called 'London Autograph' (British Library, Add. Ms. 35021). In the hands of both J. S. Bach and Anna Magdalena Bach,⁸ the London Autograph is a collection of *Auflagebögen*⁹ copied between 1739 and 1742. It can be divided into three layers according to scribes, paper type and headings:¹⁰

Layer 1 — Twelve pairs of preludes and fugues with a single type of both paper and rastrum, written in 1738–9 over a short period. It consists of pieces in commonly used keys. The movements are all entitled 'Praeludium' or 'Fuga', followed by their number in the collection's ordering.

Layer 2 — Seven pairs of preludes and fugues on several different paper types (of earliest known date 1739), with a corresponding range of rastrum types. It was written later than Layer 1, over a longer period and consists mainly of pieces in more remote keys. The preludes are all entitled 'Prelude' followed by their number in the collection; the fugues, entitled 'Fuga', are followed by the number of voices, for example 'à 3'.

Layer 3 — Two pairs of preludes and fugues on two types of paper and with two types of rastrum. It was written in 1740 at the earliest. The right-hand staff uses the treble (G₂) clef instead of the soprano (C₁) clef used in Layers 1 and 2. The titles of the two preludes refer to both the prelude and its associated fugue, while the two fugues do not share a format for their titles.

In addition to the copy in the London Autograph (Layer 3), the Fugue in A_b major is preserved in the

only other extant autograph, the Berlin Autograph. The hypothetical existence of at least one other central autograph source, which may have consisted of a collection of autograph materials, has been postulated by most studies into the compositional history of WTC II.¹¹ Three pairs of movements (in C_# minor, D major and F minor) are missing from the London Autograph, but they are all found in the extant source P 416 dating from c.1742. Although not the earliest copy, it is the only one known to have been made directly from the London Autograph¹² and is one of the most faithful copies of the London Autograph for all other surviving movements.¹³ The presence of these movements in P 416 suggests the London Autograph was a complete record of Bach's intentions at one moment in time for all pieces in WTC II. However, Bach was an inveterate reviser of his work,¹⁴ and in this respect, the London Autograph, whilst an invaluable resource, does not fully represent his final intentions for WTC II. The London Autograph itself contains a large number of revised readings, and Altnickol's 1744 copy, P 430, made whilst under Bach's tutelage and believed to be copied from the hypothetical lost autograph collection, carries readings which both pre-date (matching with extant earlier versions of some movements) and post-date the readings in the autograph.¹⁵

Earlier versions of eleven of the movements are found in several non-autograph manuscripts dating from the 1720s to the 1730s.¹⁶ In some cases these differ significantly from the versions in the London Autograph, where several movements are expanded considerably (for example, Preludes in C major and D minor) and other movements are transposed into new keys.

Transmission history

The transmission of a musical work differs from most literary traditions in reflecting both a written tradition and a parallel performance tradition. The variant readings in a source, made deliberately or in error, would undoubtedly be reflected in any performances from that source, but the transfer of information between text and performance is not a one-way process. Embellishment and interpretation of the musical text in performance may also feed back into the written tradition, as copyists may have varied the

text in ways that reflected their performance preferences. These preferences, in turn, would act as snapshots of contemporary performance practices and the variants so introduced would be transmitted as these texts were copied. The full range of variants reflects the overall reception of the work during the time period of the sources studied.

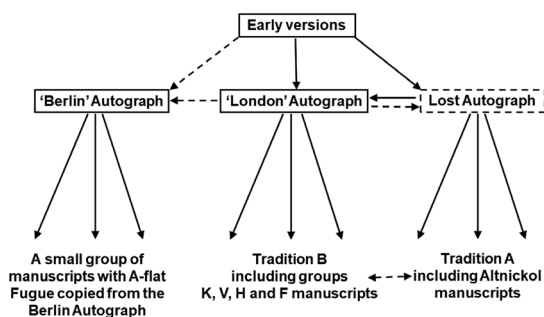
The extant sources of WTC II offer a vast, bewildering and sometimes irreconcilably contradictory record of variant readings; its transmission history is complex and elusive. Each of the three WTC II autographs (including the hypothetical autograph source/s) gave rise to distinct traditions comprising related groups of extant sources, depicted schematically in simplified format in [illustration 1](#).¹⁷ [Appendix 2](#) gives a brief description of each of the traditions and their constituent groups of manuscripts. Scholars have created stemmata or genealogical descriptions indicating the transmission histories of individual movements of the WTC II.¹⁸ The vast amount of variant data means that manual analysis of the relationships between sources has been restricted to specific significant manuscripts. In 1995 Yo Tomita published a dataset tabulating all variant readings in all extant manuscript and early printed sources for all preludes and fugues in WTC II.¹⁹ This resource consists of 16,143 variant locations (an average of 336 per movement) in up to 52 sources (the maximum number for a single movement). While it cannot realistically be analysed manually, it lends

itself to analysis by computational methods developed for phylogenetic analysis.

Phylogenetic analysis of textual and other data

Phylogenetic analyses use biological data to infer the evolutionary history of groups of organisms.²⁰ The data usually comprise DNA sequences for a particular gene, and the evolutionary relationship of the organisms can be determined from the transmission of variants introduced into the gene sequences resulting from occasional errors (mutations) that occur during copying and are inherited by subsequent generations. This inheritance of accumulated errors is, in many ways, analogous to the transmission of textual changes during repeated copying of a textual tradition by scribes. There are many similarities between the inheritance and accumulation of scribal variants and mutations in DNA molecules through successive generations. Many common types of textual change parallel those seen in DNA, such as deletions or insertions, and the formation of contaminated texts or genetic hybrids.

The inheritance of textual variants has long been exploited in the development of diagrams depicting transmission histories of textual traditions (stemmata). The first stemma is widely taken to be one generated by H. S. Collin and C. J. Schlyter (1827) indicating relationships among a set of medieval Swedish legal texts. Karl Lachmann in the first half of the 19th century, and Paul Maas in the first quarter of the 20th century, are widely credited with formalizing the principles used for generating stemmata, and the approach has been widely applied in textual criticism.²¹ However, the complexity of using variant transmission to determine relationships between a group of species or texts increases more than exponentially with an increasing number of sources or species. With the rapidly increasing availability of biological sequence data, scientists developed phylogenetic computer programs that can create trees representing evolutionary histories, using far more variant data than could be analysed manually.²² These trees are in many respects analogous to textual stemmata and, in collaborations between scientists and textual scholars, phylogenetic methods have been applied to the analysis of textual datasets such as an old Norse narrative,



1 Schematic representation of the inter-relationship between the three source traditions for the WTC II and the early versions of some of the movements. Dashed arrows indicate possible transfer of information; solid arrows indicate certain transfer.

Chaucer's *Canterbury Tales*, Dante's *Monarchia*, the German narrative *Parzival* and the Finnish legend of St Henry.²³ Phylogenetic methods have also been used in linguistic studies and with cultural artefacts such as ethnic carpet patterns where variants are transmitted from generation to generation.²⁴

The applications of stemmatic and phylogenetic analyses to textual traditions have their critics. In a landmark early 20th-century work on medieval French texts, Joseph Bédier criticized the predominance of bifurcations in stemmata.²⁵ He was refuted by Paul Maas, who determined that, with three manuscripts, bifurcating trees would be expected to outnumber any others.²⁶ More recently, Armin Hoenen *et al.* developed this mathematical approach to show that, for any number of manuscripts, there is a high probability of root bifurcations and that 'the proportions of root bifurcating stemmata observed in real collections of genealogies is close to what is mathematically predicted, with a seemingly small bias for root bifurcations'.²⁷ W. J. Phillips-Rodríguez *et al.* pointed out that statistical analysis could indicate where a multifurcation was more appropriate than multiple bifurcations.²⁸ Bédier also claimed that one set of data might be consistent with multiple stemmata, although B. J. P. Salemans argued that these might in many cases be topologically equivalent.²⁹

Others have questioned the applicability of computer-based methodologies developed for the biological sciences on the basis that genetic mutation is not directly analogous to the transmission of textual variants or errors—arguing, for example, that textual errors may be corrected, or that the same change may occur independently in multiple sources.³⁰ Howe *et al.* responded in detail to these and other criticisms.³¹ They point out that restoration of the ancestral sequence (reversion) can occur in DNA, and that identical changes in DNA sequences may occur independently in the phenomenon known as convergent change or homoplasy.

Principles of stemmatic analysis are familiar to many scholars and editors of early music;³² however, the application of phylogenetic methods in a musical context has been more restricted, although some studies have focused on the development of musical instruments or on parameters such as rhythmic patterns.³³ Some preliminary computer-based work by Penelope Rapson studied the transmission of motets

by Tallis;³⁴ however, this work from 1989 pre-dated the development of the sophisticated computer programs now available. In a 2014 article in *Early Music*, we extended the use of phylogenetic methods to a study of the extant sources of Orlando Gibbons's Prelude in G.³⁵ While the nature of the variants differs from literary textual analysis, features such as pitch, rhythmic detail and notation may be used to study the transmission of the composition.

The current article tests the use of phylogenetic analysis on a much more substantial tradition, using Yo Tomita's dataset containing detailed variant information for all movements of WTC II. As an initial case-study, we focus on two preludes, both in the hand of J. S. Bach: the Prelude in A major (BWV888) from Layer 1 of the London Autograph and the Prelude in C major (BWV870 including 870a and 870b)³⁶ from Layer 3. Both preludes are in major keys and of similar length, but they contrast strongly in their composition and transmission histories, reflecting different stages in the development of WTC II. In the London Autograph, the Prelude in A is a fair copy with only a few small corrections (*illus.2*), while the Prelude in C is a revision score with material crossed out and with numerous corrected readings, indicating the development of the piece from its earlier 17-bar form (*illus.3*).³⁷ For each prelude, the phylogenetic analysis demonstrates a grouping of sources that will be compared with the results determined by more traditional methods.

Data preparation and analysis

In Tomita's original dataset of textual variants in WTC II, each source was represented by a row and each variant was recorded as a column. The variant readings consisted of brief coded descriptions using musical notation where required (see *illus.4*).³⁸ Some columns contained information on a single simple variant, while others contained more complex information relating to several variant features.³⁹ The sections of this dataset representing the Prelude in A and the Prelude in C were expanded in consultation with images of the sources to give two tables in which each column represented a single variant feature.⁴⁰ Where there are *ante correcturam* and *post correcturam* readings in a source, separate *ante* and *post correcturam* versions of the source were prepared, each having a separate row in the variant



2 J. S. Bach, Prelude in A, BWV 888, London Autograph. British Library, Add. Ms. 35021, fol.16r (by permission of the British Library Board)

table. The data in the tables was encoded, with the top row being coded as a 'o' at each location, and with each variant reading at that location being coded by a '1', '2' etc. All sources with the same reading share the same code for that column (see [illus.5](#)).

Each variant was also categorized (for example, as 'pitch', 'repeat accidental', 'rhythm' etc.) so any category of variant could be included in or excluded from an analysis as required. Categories of variants included features that occur at a single location (for example, title, time signature, clef), and features that occur throughout the dataset (for example, various categories of accidental, pitch, staff, notation, note

value, rhythm, note/rest, tie, slur, system and page breaks, ties before and after breaks, clef change, voice and ornamentation). Decisions about which categories of variants should be excluded were made and justified prior to analysis.⁴¹ Where there was no possibility of a reading at a particular location for some sources, a '?' was recorded in the relevant column to indicate the 'missing' data. For example, [illustration 6](#) shows an extract from the variant table including a column recording the presence or absence of a tie associated with a note. Where there is no note (scored as 'o' in the relevant column), there is no possibility of a tie (scored as '?' in the



3 J. S. Bach, Prelude in C, BWV 870b, London Autograph. British Library, Add. Ms. 35021, fol.1r (by permission of the British Library Board)

relevant column). Where there is a note ('i'), there may ('o') or may not ('1') be a tie. The '?' symbol is not included in the computations performed by the phylogenetic programs. The variant tables were then used to create 'Nexus' files⁴² which can be analysed by the programs directly.

While a range of computer programs is available, the Neighbor-Net (NNet) method is particularly appropriate for a tradition where multiple affiliations between the sources may preclude the construction of an accurate bifurcating tree (or even a tree at all).⁴³ For example, a given source may be a hybrid of two or more sources, instead of being derived directly from one. Rather than creating a branching tree, NNet creates a two-dimensional network that can represent multiple affiliations between sources.

As with most phylogenetic methods, NNet does not specify any source or group as the origin ('root') of the tradition; this is important, as the origin may not be represented by any of the extant sources. NNet uses the Nexus file to create a matrix of distances between all pairs of sources (or organisms for a biological analysis). This pairwise matrix is calculated from the number of sites at which each pair of sources differ from each other, and is then used to construct a network, starting from the closest pairs of sources.

For the analyses reported here, four categories of variants were excluded:

- *Repeat and cancelling accidentals.* The conventions for indicating these accidentals within

S/N	25	26	27	28	29
Bar	4	4	4	5	5
V, bt/pos	A,3	A,4-	T;B,4	S,1-	A,1-2
Element	notation	tie	vtx	tie	notation
Spec.Loc	[g b]	♭ [c']	♭ [c] / ♭ [E A]	♭ [f']	[c' c' d' a b]
Classified	N-a,d	M-e	M-e	M-e	M-d
E: var	[♭] ud				
pitch					
P 804	E	in	ok.	in	♭ [♭] ddduu
Scheibner 4	E	in	ok.	in	♭ [♭] ddduu
P 1089	E	♭ [c']	ok.	♭ [c']	♭ [♭] ddduu
P 561	E	in	ok.	in	♭ [♭] ddduu
Memp-Prel.8	♭ [E]	in	ok.	in	♭ [♭] ddduu
N.10490	L	in	ok.	in	♭ [♭] ddduu
L: var	[♭] Lu				
pitch					
Add.35021	L	in	ok.	in	♭ [♭] Lu



4 Extract from Tomita's dataset for the Prelude in C, describing variant readings relating to specific features of the sources

- *Tie placement.* While information on the presence or absence of ties was retained, the placement of ties at a system or page break was excluded. It was felt this data would give undue

weighting to ties for sources that happened to share the same location of system or page breaks.

- *System and page break locations.* While one source (P 416) is a slavish copy of the London Autograph replicating the exact breaks, it was felt that generally this data could cause grouping of unrelated sources according to the size of the paper or the musical handwriting.

Column no.	1	2	3	Bar													16	17
				4	5	6	7	8	9	10	11	12	13	14	15			
Col. in Dataset	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
SATB	T	B	B	S	S	S	S	S	S	S	S	S	S	S	S	T	T	
Var Type	Title	time sig	clef	Tie	Note	Tie	Rhythm	Pitch	Pitch	Pitch	Pitch	RptAccd	Pitch	Pitch	RptAccd	Pitch	Tie	
P 804 ac	0	0	0	0	0	?	0	0	0	0	0	0	0	0	?	0	0	
P 804 pc	0	0	0	0	0	?	0	0	0	0	0	0	0	0	?	0	0	
Scheibner 4	0	0	0	0	0	?	0	0	0	0	0	0	0	0	?	0	0	
P 1089 ac	1	0	0	0	0	?	0	0	0	0	0	0	0	0	?	0	0	
P 1089 pc	1	0	0	0	0	?	0	0	0	0	0	0	0	0	?	0	0	
P 561	2	0	0	0	0	?	0	0	0	0	0	1	0	0	?	0	0	
Memp-Prel.8 ac	3	?	0	0	0	?	0	0	0	0	0	1	0	0	?	1	1	
Memp-Prel.8 pc	3	?	0	0	0	?	0	0	0	0	0	1	0	0	?	1	1	
N.10490	4	1	0	1	0	?	0	0	0	0	0	1	0	0	?	0	1	
Add.35021 ac	5	1	1	0	1	0	?	0	1	1	1	?	1	1	0	0	0	
Add.35021 pc	6	1	1	0	1	0	0	1	1	1	1	?	1	1	0	0	0	
Fürstenuau	?	?	?	?	?	?	0	1	1	?	?	?	?	?	?	?	?	
Go.5.312	7	1	1	0	1	1	0	1	1	1	2	?	0	1	0	0	0	
P 210 ac	8	0	1	0	1	0	0	1	1	1	2	?	0	1	0	0	0	
P 210 pc	8	0	1	0	1	0	0	1	1	1	2	?	0	1	0	0	0	
DD 70 ac	9	1	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	
DD 70 pc	9	1	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	
Am.B.57 ac	a	1	0	0	1	0	0	1	1	1	1	?	1	1	0	0	0	
Am.B.57 pc	a	1	0	0	1	0	0	1	1	1	1	?	1	1	0	0	0	
P 211 ac	b	1	0	0	1	0	0	1	1	1	1	?	1	1	0	0	0	
P 211 pc	b	1	0	0	1	0	0	1	1	1	1	?	1	1	0	0	0	
Weyse ac	c	1	0	0	1	0	0	1	1	1	1	?	1	1	0	0	0	
Weyse pc	c	1	0	0	1	0	0	1	1	1	1	?	1	1	0	0	0	
Am.B.49 ac	d	1	0	0	1	0	0	1	1	1	1	?	1	1	0	0	0	
Am.B.49 pc	d	1	0	0	1	0	0	1	1	1	1	?	1	1	0	0	0	
Weimar ac	e	1	0	0	1	0	0	1	1	1	1	?	1	1	0	0	0	
Weimar pc	e	1	0	0	1	0	0	1	1	1	1	?	1	1	0	0	0	
30386 ac	f	1	0	0	1	0	0	1	1	1	1	?	1	1	0	0	0	
30386 pc	f	1	0	0	1	0	0	1	1	1	1	?	1	1	0	0	0	

5 Extract from the variant table for the Prelude in C; sources sharing an identical variant are coded with the same number

- *Staff (and stem variants)*. Staff information was excluded due to the difficulty in determining whether notes between the staves were located on the upper or lower staves. Stem direction was largely not recorded in the variant table since there are vast numbers of stemmed notes with only two possible readings. It was hence felt that stem direction could swamp the analysis with very low-grade variant information, often influenced by scribal preference.

Results

Prelude in A BWV888

This prelude consists of 33 bars of music in three voices. Tomita's original dataset for this prelude contained 192 columns of data for 34 sources. The variant table prepared for analysis comprised 327 columns, with the number of sources increased to 53 to allow for *ante* and *post correcturam* versions of some sources. After the standard exclusions, the analysis was performed using 234 columns of data.

The network generated by NNet is shown in [illustration 7](#), where the individual source and group

names are coloured according to the groups and subgroups established by Tomita.⁴⁴ The sources fall into several clear groups, with NNet groupings corresponding closely with those specified by Tomita. The London Autograph (labelled Add. 35021) is surrounded by key sources from the H, F and K groups forming Tradition B (there are no V sources for this piece), which are known to be closely related to the London Autograph.

The Hamburg text is an important early source which is thought to be copied indirectly from the London Autograph,⁴⁵ and which may have been inherited by C. P. E. Bach who was then working in Berlin. The early editions, which derive from the intense copying activity associated with the H group of manuscripts,⁴⁶ are grouped as HEd and are located within the H group in the network. The H2 sources are believed to derive from the H1 group via a presumed lost intermediary.⁴⁷ In the network, the H2 sources mostly form a discrete group within H, but the Horn source is shifted to the HEd texts. This reflects some common notational choices shared by Horn and some of the early editions, such as a tendency to use multiple rests instead of dotted rests (for

T	B	B	S
Tie	Note	Tie	Rhythm
0	0	?	0
0	0	?	0
0	0	?	0
0	0	?	0
0	0	?	0
0	0	?	0
0	0	?	0
0	0	?	0
0	0	?	0
1	0	?	0
0	1	0	0
0	1	0	0
?	?	?	0
0	1	1	0
0	1	0	0
0	1	0	0
0	1	0	0
0	1	0	0
0	1	0	0
0	1	0	0
0	1	0	0

6 Extract from the variant table for the Prelude in C, showing use of '?' symbols for missing data

example, bar 9, alto, beat 4; bar 22, soprano, beat 2), a feature also shared by the K4 group of manuscripts.

The F group of sources comprises P 416, the direct copy of the London Autograph, and Go.S.312, which is believed to have been copied from P 416 via a lost intermediary.⁴⁸ The F sources group closely with the London Autograph and show a typical exemplar/copy grouping in which, for each pair of sources (firstly the London Autograph/P 416, and then P 416/Go.S.312), the exemplar has the shorter branch, while the copy, which carries the variants of the exemplar plus its own variants, has a longer branch.⁴⁹ The lengths of the lines in the network are a measure of the number of differences (i.e. distance) between sources.

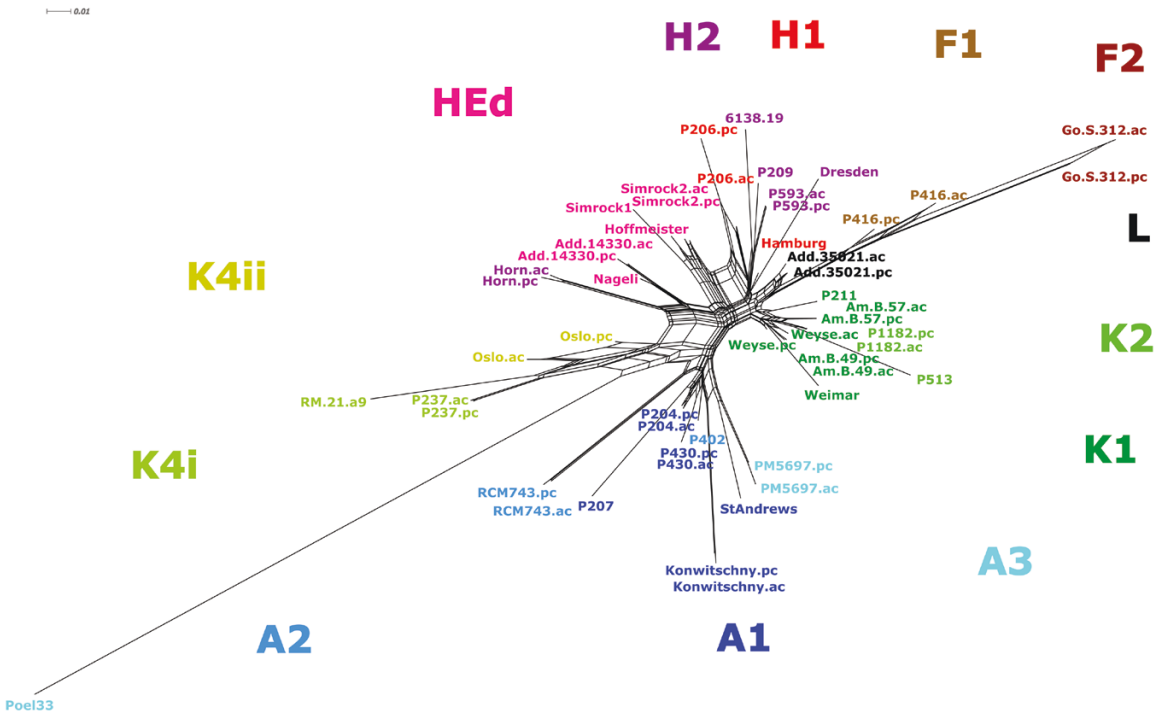
The K groups of manuscripts are linked with Bach's student Johann Philipp Kirnberger and are believed to stem from a lost intermediary, with groups K1, 2 (and 3, not represented in this prelude's

sources) stemming independently from this lost source, while K4 derives via K3. K1 (which includes Kirnberger's *Handexemplar*, Am.B.57) and K2 are considered to be closely related texts;⁵⁰ in the network, they form a discrete group close to the London Autograph and F group. K4 is less closely related and, in the absence of any K3 sources, is pulled—possibly by clef changes and the notation of rests shared with Horn, Nägeli and Add.14330, occasional absent ties in common with the highly idiosyncratic source Poel33, and some notational features shared with the A tradition generally—to group away from the other K sources and closest to the sources that form Tradition A in the network.

In illustration 7, Tradition A is represented by the A1, A2 and A3 groups of sources. The tradition is known to derive from the lost autograph (see *illus.1*),⁵¹ and the sources comprising it are separated from Tradition B in the network. Tradition A consists of sources that are copied from P 430 directly (P 204) and indirectly (Konwitschny, St Andrews and P 207). P 402 is a second copy made by Altnickol, probably from the same exemplar as P 430. RCM743 is designated as an A2 group manuscript as it shares metrical conventions with P 402, although these do not involve this prelude. However, Tomita notes that textually RCM743 is closer to the A1 group and to Poel33,⁵² and this is reflected in the NNet results. In Tomita's original study, Poel33 and PM5697 are grouped together to form A3 because of shared errors and notational readings (not involving the Prelude in A). However, this grouping is not apparent in the phylogenetic analysis (*illus.7*), owing to the large number of unique readings for the Prelude in A in Poel33 (such as the absence of a note in bar 8, alto, beat 3; pitch errors in the alto at bar 9, beat 1 and bar 27, beat 4; and more than 40 missing ties).

Prelude in C BWV870/ 870a/ 870b

This prelude consists of 34 bars (BWV870 and 870b) or 17 bars (BWV870a) in a loosely formulated four-part texture reflecting its improvisatory origins. Tomita's original dataset consisted of 381 columns for 44 sources, of which 29 sources are common to the Prelude in A. The variant table prepared for analysis comprised 933 columns, with the number of sources increased to 79 to allow for *ante* and *post correcturam* versions of some sources. The manuscript



7 NNet analysis of the variant data for the Prelude in A (with the standard exclusions). ‘Ac’ and ‘Pc’ denote *ante* and post correcturam versions respectively.

‘Fürstenau’ represents another section of the same source that formed F1 (see [Appendix 2](#)). The substantial ‘Fürstenau’ section, rediscovered in Dresden by Moritz Fürstenau in 1876, was lost in World War II, with only fragmentary information remaining; it hence is excluded from all analyses. After the standard exclusions, the analysis was performed using 710 columns of data for 78 sources. The result of NNet analysis using the standard exclusions is shown in [illustration 8](#).

The sources for the Prelude in C are clustered into three main well-separated regions. The E groups comprise early non-autograph sources written prior to the compilation of WTC II, and copies made from these early sources. All preserve the shorter 17-bar version of this piece and most date from the late 1720s to 30s. The variant table records ‘?’ readings for the early sources in the bars that are not present in these sources. Three distinct subgroups of early sources are

visible in the network. The group Ei comprises Johann Peter Kellner’s manuscript P 804, and Scheibner⁴ which is believed to use P 804 as its exemplar.⁵³ Group Eii comprises the manuscript P 1089 written by Bach’s pupil, Johann Caspar Vogler, and a considerably later copy, P 561. The Eiii manuscript N.10490 contains five preludes and fugues (in C major, D minor, E minor, F major and G major) that are partially adapted for use in WTC II. This important source is in the hand of Johann Heinrich Michel (1739–1810), who was a scribe for C. P. E. Bach in Hamburg, and it may well have been copied from a Bach autograph in the late 18th century. It has been suggested that MempPrel8 is related to P 1089,⁵⁴ although Tomita links it with N.10490 in group Eiiii, and this grouping is upheld in the network. The early sources are separated by a long branch from the London Autograph and other members of the B tradition, representing the considerable revision

A/H group that cannot generally be clearly divided into its subgroups. As with the Prelude in A, source Poel33 has the longest individual branch of the A group of sources.

Ante and post correcturam sources

In most sources for the two preludes considered here, the *ante* and *post correcturam* versions group as very close neighbours. There are very few corrections to the Prelude in A in the London Autograph and, consequently, there is little difference in branch length between the *ante* and *post correcturam* versions. For the Prelude in C, the London Autograph shows significant revisions (see [illus.3](#)). For this piece, the *post correcturam* version is believed to have been used as the exemplar for the K group of sources,³⁷ and it therefore is more representative of the text used in Tradition B for this piece, while the *ante correcturam* version is further removed, and hence has a longer branch in the network ([illus.8](#)).

Two manuscripts show a larger spacing between their *ante* and *post correcturam* versions. LM4837 is a K4ii group text carrying the Prelude in C but not the Prelude in A. In [illustration 8](#), the *post correcturam* version of LM4837 is shifted in the direction of the A and H groups, mainly reflecting their shared insertion of demisemiquavers into the opening semiquaver runs (as in the London Autograph).

The Oslo manuscript may not be of great musical significance as a source of WTC II. However, it is of some interest here, as it carries the Preludes in A and in C, and in both cases the text carries many revised readings, with the *post correcturam* version shifted considerably away from both its *ante correcturam* version and the other K4 sources in the NNet results. For the Prelude in A, the 24 revisions involve mostly the addition of ties and accidentals (although many are repeat accidentals which are excluded from this analysis). The majority of these revisions move Oslo to the reading found in both the London Autograph (and other Tradition B sources) and the A group of sources, although an occasional source may differ. There is little indication as to which tradition may have supplied the exemplar for revisions seen in the Oslo text. However, for the Prelude in C, the network shows very clear movement of the *post correcturam* version of Oslo away from Tradition B towards Tradition A ([illus.8](#)). There are over 100

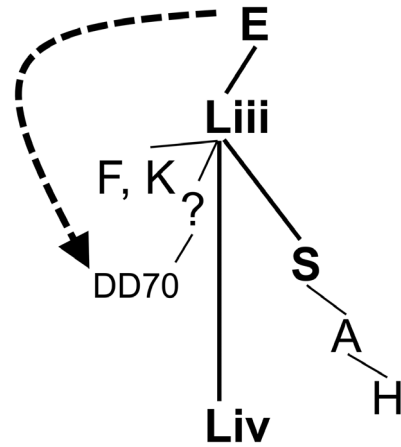
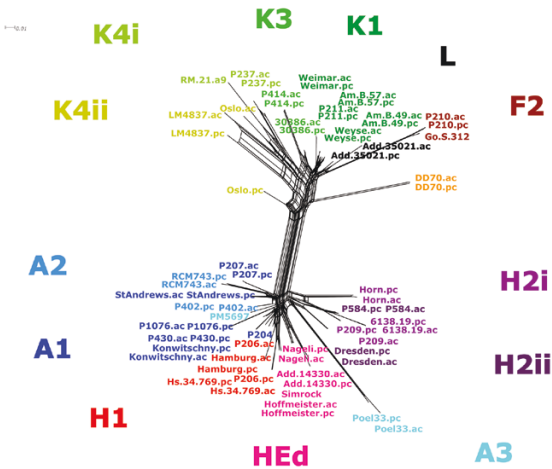
revisions, with the majority moving towards the readings found in both the London Autograph (and the majority of Tradition B) and the A and H groups of Tradition A. A few revisions, such as the removal of a bass tie in the first two beats of bar 8, give a reading closer to that of Tradition B. However, in many more cases, the *post correcturam* readings in Oslo move it clearly towards the A and H groups, for instance the insertion of rests in the tenor at bar 20 (beats 2–4) and bar 21 (beat 1), and several rhythmic, notational and pitch revisions at bar 19 (bass). The inconclusive situation for the Prelude in A, combined with the clearer shift towards Tradition A seen in the networks for the Prelude in C, may support the view that a key exemplar for the revision of Oslo may be found within Tradition A.

W. F. Bach's manuscript, DD70

The DD70 manuscript was copied by W. F. Bach in the 1740s and includes a version of the Prelude in C, but not the Prelude in A.⁵⁸ According to Tomita, it is based on the London Autograph, and may derive from a semi-finalized stage of the text.⁵⁹ In the NNet network for the Prelude in C with the standard exclusions, DD70 extends from the branch leading to the early sources ([illus.8](#)). However, the exact positioning of DD70 is finely balanced and is dependent on the precise combination of exclusions; for example, if the cancelling accidentals are not excluded from the analysis, DD70 shows linkages to both the early sources and the London Autograph.

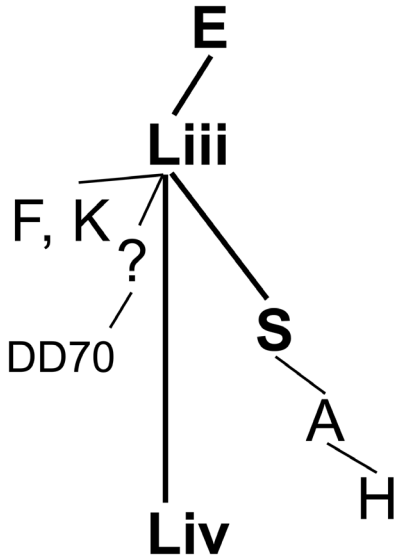
If the early sources are removed from the analysis, DD70 is positioned closest to Tradition B, but is shifted a little towards the A and H groups ([illus.9](#)). Its uncertain affiliation arises from it sharing some readings exclusively with the early sources and others exclusively with sources from either Tradition A or B. Generally, the readings that DD70 shares with either the early sources or with Tradition B are more compelling than those shared with groups A and H, which tend to be small notational differences.

DD70 appears to be linked to, but not directly part of, the main A and B traditions of WTC II. This is illustrated by a diagram of source relationships given by Tomita ([illus.10](#)), where DD70 is considered to derive from one of the later stages of development of the London Autograph.⁶⁰ The affiliation between DD70 and the early sources cannot be fully



9 NNet analysis of the variant data for the Prelude in C (with the standard exclusions, and the early sources deleted from the analysis)

11 Indication of information transfer for the Prelude in C from the early sources to manuscript DD70



10 Diagram of possible source relationships for the Prelude in C

explained by this diagram and requires a transfer of information from the early sources to DD70 as indicated in illustration 11.

Overall, DD70 shows a complex pattern of shared and unique readings that often shift between the main traditions. There are some striking agreements, such as the unique combination of

motivic ideas in the second half of bar 1 (see [ex.1](#)). At beat 3, DD70 shares a reading of *e'*, *c''*, *g'*, *b'* with Tradition B only (here represented by the London Autograph and Group F). Then at beat 4 it agrees solely with the early sources (Group E), with a reading of *c'*, *b'*, *a'*, *g'*. W. F. Bach's performance choices are also seen in a number of unique readings (see [ex.2](#)). In the second half of bar 3, the combination of pitches is unique; in bars 6 and 21, the ornamentation is altered by translating the right hand's written-out mordent into the left hand; in bars 9, 19 and 24, the harmony is intensified with additional chromatic notes (although the variant in bar 19 was subsequently erased); and in bars 15–16, the syncopated bass part forms a *brisé* texture with the alto part, then the tenor.

There are several possible explanations for this complex pattern of shared and unique readings. The agreements may be a result of W. F. Bach having multiple exemplars available or, far more likely given his stature as a musician, he may have developed his own version of the piece, possibly remembering readings from multiple manuscripts to which he may have had access over the years.⁶¹ Another possibility is tantalizingly suggested by the only other piece from WTC II included in DD70. Like the Prelude in C, the Fugue in A^b is a Layer 3 piece in the London Autograph, and DD70 is one of the small group of manuscripts in which the

Fugue in A \flat is believed to be copied from the Berlin Autograph (see [illus.1](#)).⁶² The Berlin autograph is part of the miscellany P 274 assembled in the mid 19th century, which includes the Fugue in A \flat from WTC II alongside unrelated works by Bach.⁶³ It is conceivable that the version of the Prelude in C found in DD70 was copied from a lost version linked with the copy of the Fugue in A \flat now found in the Berlin Autograph.⁶⁴ An affiliation between DD70 and the Berlin Autograph for the Prelude in C cannot be indicated directly from the analyses presented here, as the prelude is not present in the Berlin Autograph. However, such an interpretation would be consistent with the separation of DD70 from the other main traditions in the networks and its generally unsettled affiliations.

Discussion

Tomita's 1995 dataset of WTC II readings comprises a vast quantity of information on variants that cannot possibly be analysed manually. Applying phylogenetic analysis to this dataset enabled a visualization of the relationships between the various groups of sources, in a manner analogous to looking at the evolutionary relationships between species in biological analyses. While a manual analysis, taking account of textual and codicological features such as paper and binding, may be ideal for a small dataset of musical texts, computational analysis is the most effective tool for dealing with a large amount of data. It yields provisional indications of relationships that can then inform a full assessment which may include other parameters such as codicology.

As already mentioned, the Preludes in A and in C contrast strongly in their compositional and transmission histories. The Prelude in A represents a piece which Bach saw little need to revise. The lack of revision is reflected in the NNet network, where the distances between groups of sources is little greater than the distances between sources within the same group. The situation for the Prelude in C is entirely different, with a succession of revisions already within the London Autograph (see [illus.3](#)). The significant distances between the groups of sources in the network for this prelude are consistent with the process of continued revision, as Bach moved away from the version(s) in the early sources, through the composing score of the London Autograph and its copies, to the further revision seen in the Altnickol group.

In the phylogenetic analysis of the two preludes from WTC II, the grouping of sources by NNet agreed well with the main groupings established by more traditional methods, although the data did not always provide an adequate signal to allow reliable distinction between closely related subgroups. The analyses suggested possible exemplars for the revisions found in the Oslo manuscript and highlighted the independent nature of the manuscript DD70, which may represent a different line of transmission. The DD70 version of the Prelude in C contains an unusual mixture of readings and unique variants, perhaps reflecting compilation from memory rather than by direct copying of a physical exemplar. Many of these variants make unexpected changes at a fundamental compositional level and engage with the implications of the prelude's complex compositional history. By highlighting the broad affiliations of DD70, the analysis highlights its importance as a witness of W. F. Bach's performance choices concerning his father's music and the inherent potential of the expanded Prelude in C for further musical development. In addition to changes of harmony and rhythm, W. F. Bach clarified some of the ambiguity that arises when notating broken-chord figuration polyphonically. Such ambiguities can be interpreted or resolved in many different ways, and hence pose both interpretative problems and extempore opportunities for performers. To have W. F. Bach's choices, for example the unique reading of the over-holding of the alto's *f* and *d'* in bar 33 ([ex.3](#)), is particularly informative for the interpretation of J. S. Bach's music and for a better understanding of 18th-century performance practice.

Although the dataset of variants was time consuming and exacting to prepare, phylogenetic analysis offers flexible and speedy analysis of large amounts of complex material. It allows a focus on particular sections of a tradition, or experiments in the effects of omitting selected sources or types of variant. It also raises the possibility of analysing categories of variants that may be particularly associated with performance, to gain insight into the transmission of a composition from the perspective of performers. 'Lost' historical performances would arguably be represented as internal nodes in the underlying phylogeny of the tradition. It is important that the source material is consistent with the assumptions of the phylogenetic approach,

Ex.1 J. S. Bach, Prelude in C, BWV870, bar 1, right hand; DD70 compared with other sources

DD70



F group



E group (BWV870a)



London Autograph (BWV870b)



P 430 (BWV870)



Ex.2 J. S. Bach, Prelude in C, BWV870: a selection of unique readings in DD70

bar 3 (beats 3–4)



bar 6 (beat 4)



bar 21 (beat 2)



bar 9 (beats 3–4)



bar 24 (beats 1–2)

bar 19 (beats 3–4 *ante correcturam* reading)

bar 24 (beats 3–4)



bars 15–16



Ex.3 J. S. Bach, Prelude in C, BWV870, bar 33 (beat 4)–34, right hand: unique reading in DD70



for example that most of the changes made in one copy are inherited in subsequent copies, and that the same change rarely occurs independently in multiple copies. Notwithstanding this requirement, the consistency between the conclusions drawn here and traditional scholarship demonstrates the potential and power of the phylogenetic approach.

APPENDIX 1

ALPHABETICAL LIST OF MANUSCRIPT SOURCES FOR THE *WELL-TEMPERED CLAVIER II*

The descriptive form of library sigla and shelfmark follows the source database of Bach Digital (www.bach-digital.de), where further information about the source can be found

Abbreviated form used	Source name as given in Bach Digital
6138.19	D-Bhm 6138/19, Faszikel 2 (Prelude in C); Faszikel 11 (Prelude in A)
30386	D-B Mus.ms. 30386, Faszikel 3
Add. 14330	GB-Lbl Add. Ms. 14330
Add. 35021	(see London Autograph)
Am.B.49	D-B Am.B 49/II
Am.B.57	D-B Am.B 57, Faszikel 2
Berlin Autograph	D-B Mus. ms. Bach P 274, Faszikel 4
DD70	I-Bc DD 70
Dresden	D-Dl Mus. 2405-T-7,2
Fürstenuau	Verschollen Teilabschrift Wohltemperiertes Klavier II [früher M. Fürstenuau]
Go.S.312	D-LEb Go. S. 312
Hamburg	D-Hs M B/1974, Faszikel 2
Horn	D-LEb Rara Ib, 168; olim: D-Gb Ms. Horn
Hs.34.769	A-Wn Mus. Hs. 34769
Konwitschny	Verschollen BWV 870–893, F. Konwitschny
LM4837	US-NH LM 4837 [Ma21.Y11.B12]
London Autograph	GB-Lbl Add. Ms. 35021
MempPrel8	D-LEb Peters Ms. 8, Faszikel 18 (Depositum im Bach-Archiv); <i>olim</i> : D-LEm, Sammlung Mempel-Preller, Ms.8
N.10490	D-B N.Mus.ms. 10490
Oslo	N-Onm Mus. ms. a 217:786
P 204	D-B Mus.ms. Bach P 204
P 206	D-B Mus.ms. Bach P 206
P 207	D-B Mus.ms. Bach P 207
P 209	D-B Mus.ms. Bach P 209, Faszikel 21 (Prelude in C)
P 210	D-B Mus.ms. Bach P 210
P 211	D-B Mus.ms. Bach P 211
P 237	D-B Mus.ms. Bach P 237, Faszikel 2
P 274	(see Berlin Autograph)
P 402	D-B Mus.ms. Bach P 402, Faszikel 2
P 414	D-B Mus.ms. Bach P 414
P 416	D-B Mus.ms. Bach P 416, Faszikel 4
P 430	D-B Mus. ms. Bach P 430
P 513	D-B Mus.ms. Bach P 513
P 561	D-B Mus.ms. Bach P 561
P 584	D-B Mus.ms. Bach P 584
P 593	D-B Mus.ms. Bach P 593
P 804	D-B Mus.ms. Bach P 804, Faszikel 38
P 1076	D-B Mus.ms. Bach P 1076
P 1089	D-B Mus.ms. Bach P 1089
P 1182	D-B Mus.ms. Bach P 1182
PM5697	D-LEb Peters PM 5697 (Depositum im Bach-Archiv); <i>olim</i> : D-LEb Peters 14314c

Poel33	D-LEm Poel. mus. Ms. 33,2
RCM743	GB-Lcm Ms. 743
RM.21.a9	GB-Lbl R. M. 21. a. 9
Scheibner4	D-LEb Peters Ms. 4, Faszikel 2 (Depositum im Bach-Archiv); olim: D-LEm Sammlung Scheibner, Ms. 4
St Andrews	GB-SA ms M24. B2
Weimar	D-WRgs Goethe Notensammlung; GSA 32/430
Weyse	DK-Kk mu 9412.0982 Weyses Samling (C I, 105)

APPENDIX 2

BRIEF DESCRIPTION OF THE SOURCE TRADITIONS AND MANUSCRIPT GROUPINGS FOR THE *WELL-TEMPERED CLAVIER II*

Tradition Provenance

E	Early model, originating in the pieces written in the 1720s. Eleven pre-WTC II pieces are known to have existed, namely the Preludes and Fugues in C major, C \sharp major, D minor and G major; and the Fugues in C minor, E \flat major and A \flat major.
B	Bach Tradition, stemming from the London Autograph (L).
A	Altnickol Tradition, stemming from another set of Bach's autograph (S) which is no longer extant, but is attested to in surviving sources.

Ms. Group Features of manuscript group

E	Early models, P 804 , P 1089 , P 595 , P 226 , etc. Copies from the source Bach originally composed in c.1720–38. Where applicable, they are divided into several subgroups according to their textual chronology.
L	London Autograph, Add. Ms. 35021 , compiled in c.1739–42. The Preludes in F major, G minor, C \sharp minor, D major and F minor are missing from the collection.
V	Vienna manuscripts of 24 Fugues only, Q 10782 (A-Wgm VII 8802 (Q 10782)), Q 11731 (A-Wgm VII 8802 (Q 11731)), Stockholm (S-Smf MMS 242) and Berea (US-BER Ma. M5 (Kenney 543)). Their common model originated in c.1741. Its text is heavily corrupt and thought to have gone through systematic revision in many layers.
H	The model which became the basis for H1 = Hamburg with its copy, P 206 , and H2 consisting of P 209 and Dresden , originated in c.1741. Apart from the Preludes and Fugues in C major and A \flat major, it is derived from L. Except for P 209 , they all have their origin in Dresden . In many instances, H2 gives an earlier text, suggesting an independent path. Apart from P 209 , each source contains a complete set.
F	Direct copy made from L in 1742, now consists of Fürstenau (lost), P 416 , GB-Lbl Add. Ms. 38068 and Chicago (US-Cn Case MS 6A 72). They are referred to as F1 when they are distinguished from their derivants. Their text is characterized by readings newer than those in H but, at the same time, the scribe often interprets the amended symbols in a different way from what Bach intended. From it stems a group F2 (Go.S.312 and P 210) which inherits a confused collection of readings. Someone made later amendments to the text of F based on A.
K	Kirnberger circle manuscripts, formed from assembled copies deriving from various stages of L.
Bn	Berlin Autograph, P 274 . Made in c.1743–46. Fugue in A \flat major only.
S	A complete autograph manuscript of WTC II, thought to have existed but now lost. It first existed as a sketch which Bach worked on up to 1739 and used as an exemplar to make L. The other evidence in the copies of S also indicates that it was a collection of revision scores. When L was completed in 1742, Bach came back to S to fill in the missing movements.
A1	Altnickol's first copy, P 430 and the copies stemming from it, consisting of St Andrews , P 204 , Konwischny , P 1076 , P 207 and others. P 430 was made under Bach's instruction in 1744. Bach is thought to have made later amendments directly onto this copy. There are other later amendments of spurious origin by several hands. One of these is F. A. Grasnick (c.1800–77) who compared A1 with K1 , edited A1 , and subsequently wrote P 1146 , in which he listed down the noteworthy variant readings he found in K1 .

- A2** Altnickol's second copy, **P 402**, made in or before 1755. It contains some later as well as earlier readings than **A1**. Altnickol did not use **P 430** itself for **A2**, but used a source closely related to that used for his earlier copy. **RCM 743** also belongs to this group.
- A3** **Poel.33,2** and **PM 5697**. A branch stemming from **S** independently of **A1** or **A2**. It contains valuable information that is not found in Altnickol's manuscripts. The former is of particular interest, for it was copied in 1767 (?) by Johann Christoph Georg Bach (1747–1814), Bach's distant cousin in Ohrdruf. The credibility of the information suffers much from careless copying practices, however.
- K1** **Am.B.57**, so-called 'Kirnbergers Handexemplar', **Weyse, Am.B.49**, **Weimar** and **P 211**. Mainstream manuscripts of Kirnberger, stemming from various stages of **L** from the earliest to the latest. This group of manuscripts is thought to have been compared with **L** in Berlin. The most probable date is 1774 or soon after when W. F. Bach, in whose possession it was, moved to Berlin.
- K2** **P 814**, **P 1182**, **P 513**, **RCM 26** and **Cambridge** (GB-Cfm MU. Ms. 161A). Originated in the model for **K1**, partly being attested by its more faithful nature to the original, and partly characterized by unique variant readings that are perhaps not derived from Bach.
- K3** **Ms.30386**, **P 414**. Stemming from the earliest branch of **K**, characterized by containing fewer variants and errors than **K1** but having its own unique ones.
- K4** **P 237**, **RM.21.a.9**, **Oslo** and **LM 4837**. Farthest group from **K1**, characterized by corruptions of the text and by many unique variant readings, occasionally shared with **K3**. While the first two manuscripts contain a complete set of WTC II, the last, the text of which is severely corrupted, contains pieces from the Prelude in C major to the Fugue in F minor, and from the Prelude in G major to the Fugue in G minor (up to bar 20).

Heather Windram is a researcher in the Department of Biochemistry at the University of Cambridge. With Christopher Howe and others, she has published widely on the application of phylogenetic methods to non-biological data, including both literary and musical textual traditions. With Terence Charlston and Andrew Woolley, she co-edits an ongoing series of facsimile editions of English Keyboard Music c.1650–c.1700. hfw30@cam.ac.uk

Terence Charlston is an internationally acknowledged early keyboard performer. He has initiated many pioneering concerts and recording projects including editions and recordings of Matthew Locke, Carlo Ignazio Monza, Albertus Bryne, William Byrd's My Ladye Nevells Booke, and French clavichord music. He teaches harpsichord at the Royal College of Music, London, where he is Chair of Historical Keyboard Instruments and was awarded the FRCM in 2020. www.charlston.co.uk

Yo Tomita is Professor of Musicology at Queen's University Belfast, and Senior Fellow of the Bach-Archiv Leipzig. He received his doctorate from the University of Leeds in 1991 with a dissertation on the sources of Bach's Well-Tempered Clavier II. He has published widely on topics ranging from Bach's compositional choices and decisions as manifested in his autographs to the reception history of Bach's music in the late 18th and early 19th centuries. y.tomita@qub.ac.uk

Christopher Howe is Professor of Plant and Microbial Biochemistry at the University of Cambridge, and has a particular interest in the evolution of proteins involved in photosynthesis. He has collaborated for many years with Heather Windram and others on the interdisciplinary application of biological methods of evolutionary analysis ('phylogenetics') to non-biological data. ch26@cam.ac.uk

¹ Y. Tomita, J. S. Bach's 'Das Wohltemperierte Clavier II': a critical commentary. Vol.ii. All the extant manuscripts (Leeds, 1995), p.vii.

² Staatsbibliothek zu Berlin (D-B): Mus.ms. Bach P 415. One leaf is now missing; it formerly was located between folios 25 and 26, and

contained the musical text of Fugue 13 in F# major and the first seven bars and three beats of Prelude 14 in F# minor.

³ In translation, the title-page reads: 'The Well-Tempered Clavier, or, Preludes and Fugues through all the tones and semitones both as regards the *tertia major* or *Ut Re Mi* and as concerns the *tertia minor* or *Re Mi Fa*. For the Use and Profit of the Musical Youth Desirous of Learning as well as for the Pastime of those Already Skilled in this Study, drawn up and written by Johann Sebastian Bach. p.t. Capellmeister to His Serene Highness, the Prince of Anhalt-Cöthen, etc. and Director of His Chamber Music. Anno 1722.' *The new Bach reader: a life of Johann Sebastian Bach in letters and documents*, ed. H. T. David and A. Mendel, rev. C. Wolff (New York, 1998), p.97.

⁴ The title-page in Altnickol's 1744 copy reads: 'Des Wohltemperirten Claviers, Zweyter Theil, bestehend in Præludien und Fugen durch alle Tone und Semitonien verfertigt von Johann Sebastian Bach, Königlich Pohlnisch und Churfurstl. Sächs. Hoff Compositeur Capellmeister und *Directore Chori Musici* In Leipzig.' English translation based on Y. Tomita, CD liner notes for *Das Wohltemperierte Klavier II* (Masaaki Suzuki, harpsichord, BIS-CD-1513/14) 2008, which is also available at <https://www.qub.ac.uk/tomita/essay/Tomita-wtc2.pdf> (accessed 13 July 2021). The authors gratefully acknowledge the kind assistance of David Charlston with this and other points of translation.

⁵ Y. Tomita, CD liner notes for *Das Wohltemperierte Klavier II*, p.10.

⁶ R. D. P. Jones, *The creative development of Johann Sebastian Bach, vol.ii: 1717–1750: music to delight the spirit* (New York, 2013), pp.335–46.

⁷ Y. Tomita, 'The implication of Bach's introduction of new fugal techniques and procedures in *The Well-Tempered Clavier* Book Two', *Understanding Bach*, vi (2011), pp.35–50.

⁸ W. Emery, 'The London Autograph of "The Forty-Eight"', *Music & Letters*, xxxiv/2 (1953), pp.106–23.

⁹ *Auflagebögen* are bifolio manuscript pages whose layout is ideal for performance from an opened individual sheet. The music cannot be read in

the correct sequence if the sheets are bound at the spine, but a single turn of the complete leaf allows for correct performance. See Y. Tomita, 'Manuscripts', in *The Routledge research companion to Johann Sebastian Bach*, ed. R. A. Leaver (Abingdon, 2017), pp.47–88, at p.63 n.53.

¹⁰ Y. Kobayashi, 'Zur Chronologie der Spätwerke Johann Sebastian Bachs: Kompositions- und Aufführungstätigkeit von 1736 bis 1750', *Bach-Jahrbuch*, lxxiv (1988), pp.7–72, at pp.45–6.

¹¹ W. Breckoff, 'Zur Entstehungsgeschichte des zweiten Wohltemperierten Klaviers von Johann Sebastian Bach' (PhD diss., Tübingen Universität, 1965), p.57, and W. Dehnhard, *J. S. Bach: Das Wohltemperierte Klavier II* (Vienna, 1983/R2001), pp.xxi–xxii.

¹² Marks in the London Autograph in the hand of P 416's scribe confirm that P 416 is a direct copy of the autograph. See Dehnhard, *J. S. Bach: Das Wohltemperierte Klavier II*, p.xxii.

¹³ J. A. Brokaw II, 'Recent research on the sources and genesis of Bach's *Well-Tempered Clavier*, Book II', *Bach*, xvi/3 (1985), pp.17–35, at p.21.

¹⁴ G. Stauffer, 'Bach as reviser of his own keyboard works', *Early Music*, xiii (1985), pp.185–98.

¹⁵ Breckoff, 'Zur Entstehungsgeschichte', p.57; Dehnhard, *J. S. Bach: Das Wohltemperierte Klavier II*, pp.xxii–xxiii; R. D. P. Jones, 'Stages in the development of Bach's *The Well-Tempered Clavier II*', *The Musical Times*, cxxxii (1991), pp.441–6.

¹⁶ Tomita, *J. S. Bach's 'Das Wohltemperierte Clavier II': a critical commentary. Vol.ii*, p.ix.

¹⁷ Based on information from Y. Tomita, *J. S. Bach's 'Das Wohltemperierte Clavier II': a critical commentary. Vol.i: autograph manuscripts* (Leeds, 1993), pp.6–8.

¹⁸ Dehnhard, *J. S. Bach: Das Wohltemperierte Klavier II*, pp.xxiii–xxxii. Tomita, *J. S. Bach's 'Das Wohltemperierte Clavier II': a critical commentary. Vol.i*, pp.17–142.

¹⁹ For detailed information on each variant, see Tomita, *J. S. Bach's 'Das*

Wohltemperierte Clavier II': a critical commentary. Vol.ii, pp.1–1001.

²⁰ A. W. F. Edwards, 'Statistical methods for evolutionary trees', *Genetics*, clxxxiii (2009), pp.5–12, gives an overview of the development of phylogenetic methods for use with biological data.

²¹ P. Maas, *Textkritik* (Leipzig and Berlin, 1927); trans. B. Flower as *Textual criticism* (Oxford, 1958).

²² R. D. M. Page and E. C. Holmes, *Molecular evolution: a phylogenetic approach* (Oxford, 1998) provides an overview of many commonly used phylogenetic methods.

²³ P. M. W. Robinson and R. J. O'Hara, 'Cladistic analysis of an Old Norse manuscript tradition', *Research in Humanities Computing*, iv (1996), pp.115–37; A. C. Barbrook, C. J. Howe, N. Blake and P. Robinson, 'The phylogeny of *The Canterbury Tales*', *Nature*, cccxciv (1998), p.839; 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; 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; T. Roos, T. Heikkilä and P. Myllymäki, 'A compression-based method for stemmatic analysis', in *17th European conference on artificial intelligence: Proceedings* (Amsterdam, 2006), pp.805–6.

²⁴ 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; J. Tehrani and M. Collard, 'Investigating cultural evolution through biological phylogenetic analysis of Turkmen textiles', *Journal of Anthropological Archaeology*, xxi (2002), pp.443–63.

²⁵ J. Bédier, *Le lai de l'ombre par Jean Renart* (Paris, 1913), p.xxiv.

²⁶ P. Maas, 'Leitfehler und Stemmatische Typen', *Byzantinische Zeitschrift*, xxxvii (1937), pp.289–94.

²⁷ A. Hoenen, S. Eger and R. Gehrke, 'How many stemmata with root degree

k?’, in *Proceedings of the 15th meeting on the mathematics of language* (London, 2017), pp.11–21.

²⁸ W. J. Phillips-Rodriguez, C. J. Howe and H. F. Windram, ‘Some considerations about bifurcation in diagrams representing the written tradition of the Mahabharata’, *Vienna Journal of South Asian Studies*, lii–liii (2010), pp.29–43.

²⁹ B. J. P. Salemans, ‘Building stemmas with the computer in a cladistic, neo-Lachmannian way’ (PhD diss., University of Nijmegen, 2000), pp.34–5.

³⁰ R. Hanna, ‘The application of thought to textual criticism in all modes—with apologies to A. E. Housman’, *Studies in Bibliography*, liii (2000), pp.163–72.

³¹ C. J. Howe, R. Connolly and H. F. Windram, ‘Responding to criticisms of phylogenetic methods in stemmatology’, *Studies in English Literature*, lii (2012), pp.51–67.

³² 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 (1992), pp.373–427.

³³ I. Tëmkin and N. Eldredge, ‘Phylogenetics and material cultural evolution’, *Current Anthropology*, xlviii (2007), pp.146–53; C. Guastavino, F. Gómez, G. Toussaint, F. Marandola and E. Gómez, ‘Measuring similarity between flamenco rhythmic patterns’, *Journal of New Music Research*, xxxviii (2009), pp.129–38.

³⁴ P. Rapson, *A technique for identifying textual errors and its application to the sources of music by Thomas Tallis* (New York, 1989).

³⁵ H. F. Windram, T. Charlston and C. J. Howe, ‘A phylogenetic analysis of Orlando Gibbons’s Prelude in G’, *Early Music*, xlii (2014), pp.515–28.

³⁶ In the 1998 version of the Bach-Werke-Verzeichnis, BWV870 generally

refers to the text of Tradition A, while BWV870a refers to the version found in the early sources and BWV870b to the version found in the London Autograph and its related sources. In the 2022 revision of the Bach-Werke-Verzeichnis, BWV870 refers to the work encompassing all its variants, while the major variants are described as follows: BWV870.1: early version (formerly BWV870a); BWV870.2: London autograph (formerly BWV870b); BWV870.3: Altnickol version.

³⁷ The crossed-out version of the transition material used to move from the shorter to the longer form of the piece is not found in any other extant source, and was not included in the original dataset or the current analyses.

³⁸ The dataset was prepared using the Bach font developed by Tomita (<https://www.qub.ac.uk/tomita/bachfont/>).

³⁹ Fingering is present in very few sources and, as it is distinct from the text of the music and could potentially have been added on a later occasion, it was excluded from the original dataset and the variant table.

⁴⁰ In addition to the lost Fürstenau manuscript, images for Hs.34.769 (Prelude in C) and 6138.19 (Prelude in A) were not available for consultation; however, these latter two manuscripts were fully represented in the original dataset.

⁴¹ This decision process is not specific for musical data. With textual traditions, decisions must be made about (for example) spelling variants, which are often excluded as they may group sources by dialect and scribal preference rather than by transmission history.

⁴² The Nexus file format is described in D. R. Maddison, D. L. Swofford and W. P. Maddison, ‘NEXUS: An extensible file format for systematic information’, *Systematic Biology*, xlvii (1997), pp.590–621.

⁴³ D. Bryant and V. Moulton, ‘Neighbor-Net: an agglomerative method for the construction of phylogenetic networks’, *Molecular*

Biology and Evolution, xxi (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 (2006), pp.254–67.

⁴⁴ Tomita, J. S. *Bach’s ‘Das Wohltemperierte Clavier II’: a critical commentary. Vol.ii*, p.774.

⁴⁵ Tomita, J. S. *Bach’s ‘Das Wohltemperierte Clavier II’: a critical commentary. Vol.ii*, p.x. A manuscript copy by S. Hering, dated ‘1742’ on the title-page, had been missing since 1924 and resurfaced too late to be included in these analyses. It has long been thought to shed some light on the transmission history of the Hamburg manuscript. However, Michael Maul, who first reported his discovery in the round table ‘Current source studies in Bach research’ at the International Musicological Society’s conference in Tokyo on 23 March 2017, notes that the musical text of the Hering source stems from Staatsbibliothek zu Berlin, Am.B.57/2 (the so-called Kirnberger’s personal copy in the hand of Johann Nicolaus Schober, formerly known as ‘Anon. 402’, dated between 1758 and 1763), possibly copied directly from its *post correcturam* state.

⁴⁶ Tomita, J. S. *Bach’s ‘Das Wohltemperierte Clavier II’: a critical commentary. Vol.ii*, pp.ix–x.

⁴⁷ Tomita, J. S. *Bach’s ‘Das Wohltemperierte Clavier II’: a critical commentary. Vol.ii*, p.x.

⁴⁸ Tomita, J. S. *Bach’s ‘Das Wohltemperierte Clavier II’: a critical commentary. Vol.ii*, p.xii.

⁴⁹ 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.

⁵⁰ Tomita, J. S. *Bach’s ‘Das Wohltemperierte Clavier II’: a critical commentary. Vol.ii*, p.xii.

⁵¹ Tomita, J. S. Bach's 'Das Wohltemperierte Clavier II': a critical commentary. Vol.ii, p.xiii.

⁵² Tomita, J. S. Bach's 'Das Wohltemperierte Clavier II': a critical commentary. Vol.ii, p.xv.

⁵³ Information about the early sources is taken mainly from Brokaw, 'Recent research'.

⁵⁴ Brokaw, 'Recent research', p.29, credits this observation to Breckoff, 'Zur Entstehungsgeschichte des zweiten Wohltemperierten Klaviers'.

⁵⁵ J. A. Brokaw II, 'The genesis of the Prelude in C Major, BWV870', in *Bach Studies*, ed. D. Franklin (Cambridge, 1989), pp.225–39, considers in detail its revisions and development.

⁵⁶ Tomita, J. S. Bach's 'Das Wohltemperierte Clavier II': a critical commentary. Vol.ii, p.1.

⁵⁷ Tomita, J. S. Bach's 'Das Wohltemperierte Clavier II': a critical commentary. Vol.ii, p.xii.

⁵⁸ On the dating of DD70, see Jones, 'Stages in the development', p.442. According to Alfred Dürr (*Neue Bach Ausgabe, Kritischer Bericht*, v/6.2, p.98), Peter Wollny suggests the period c.1735–45 from the characteristics of W. F. Bach's handwriting.

⁵⁹ Tomita, J. S. Bach's 'Das Wohltemperierte Clavier II': a critical commentary. Vol.i, p.18; also Vol.ii, p.1.

⁶⁰ Tomita, J. S. Bach's 'Das Wohltemperierte Clavier II': a critical commentary. Vol.i, p.18.

⁶¹ It is also possible that the heavily revised state of the score encouraged W. F. Bach to create his own version of this prelude.

⁶² Tomita, J. S. Bach's 'Das Wohltemperierte Clavier II': a critical commentary. Vol.ii, p.689.

⁶³ *Neue Bach Ausgabe. Kritischer Bericht*, iv/5 + 6, p.716, v/9.2, pp.247, 261, 274; also Y. Tomita, 'Revisiting the reception of J. S. Bach's *Well-Tempered Clavier II* in pre-Classical Vienna', in *A musical offering: essays in honour of Gerard Gillen*, ed. K. Houston and H. White (Dublin, 2017), pp.318–59, at pp.327–8.

⁶⁴ Jones, 'Stages in the development' (p.442) hints at this relationship: 'It is probably no mere chance, therefore, that F2 [DD70 in the current article], copied from β [the unknown DD70 exemplar] in the C major prelude, was copied from F1 [the Berlin Autograph] in the A♭ fugue'.

Heather F. Windram, Terence
Charlston, Yo Tomita,
Christopher J. Howe

A phylogenetic analysis of two preludes from J. S. Bach's *Well- Tempered Clavier II*

J. S. Bach's *Well-Tempered Clavier II* is well known for the complexity of its source situation, and its vast array of variant readings. The current article uses techniques of phylogenetic analysis, developed in the biological sciences, to deepen our understanding of the complex relationships between the primary sources. The computer algorithm NeighborNet is

used to analyse data comprising the textual variants for the Prelude in A, BWV888 and the Prelude in C, BWV870. The resultant grouping of sources reflects the differences in revision practice between the two preludes. While Bach saw little need to revise the Prelude in A, the Prelude in C underwent a process of continued revision that can be discerned in the results of the phylogenetic analysis. The analyses also highlight the uncertain relationship of the manuscript DD70 with the other sources of the Prelude in C and the implications for 18th-century performance practice.

Keywords: phylogenetics; Johann Sebastian Bach; *Well-Tempered Clavier II*; evolutionary algorithms; NeighborNet; transmission history; performance practice