



Aligning the Odes: A Character-Level Parallel Dataset for the Anhui University and Fùyáng Bamboo-Slip Manuscripts

DATA PAPER

JULIEN BALEY

JĪNQÍ YĪNG 應金琦

**Author affiliations can be found in the back matter of this article*

ju[ubiquity press

ABSTRACT

This dataset aligns graph-level transcriptions of three witnesses to the *Book of Odes*: the received *Máoshī* 毛詩, the Warring States so-called ‘Anhui University’ (Āndà 安大) bamboo manuscript, and the Western Han Fùyáng 阜陽 bamboo manuscript. The release represents each witness in Unicode JSON, uses Ideographic Description Sequences for graphs without suitable encoded characters, and records explicit one-to-one, one-to-many, and many-to-many correspondences. Separate mappings connect transcription positions to physical locations in the source editions and those locations to images of the manuscript characters. JSON Schemas, integrity tests, a static read-only viewer, and reviewed crop recipes support validation and reproducibility. The corpus makes manuscript variation available both as structured data and through a visual interface in which readers can inspect aligned passages, graph images, and their positions on reconstructed manuscript strips. Its format can accommodate later manuscript publications and competing transcriptions of the same physical witness. It turns manuscript and transcription variation into an inspectable, reusable research infrastructure for early Chinese textual transmission, palaeography, historical phonology, and digital philology.

CORRESPONDING AUTHOR:

Julien Baley

School of Languages, Cultures,
and Linguistics, SOAS
University of London, UK

julien.baley@gmail.com

KEYWORDS:

Book of Odes; *Shijing*; bamboo
manuscripts; parallel
corpus; textual variation;
digital philology; Ideographic
Description Sequences

TO CITE THIS ARTICLE:

Baley, J., & Yīng. J. 應金琦.
(2026). Aligning the Odes:
A Character-Level Parallel
Dataset for the Anhui
University and Fùyáng
Bamboo-Slip Manuscripts.
*Journal of Open Humanities
Data*, 12: 138. DOI: [https://
doi.org/10.5334/johd.644](https://doi.org/10.5334/johd.644)

REPOSITORY LOCATION

<https://doi.org/10.5281/zenodo.22143305> (version 1.0.4).

CONTEXT

The *Shijing* 詩經, conventionally translated as the *Book of Odes*, *Book of Songs*, or *Classic of Poetry*, is an anthology of 305 poems transmitted as one of the canonical works of early China (Waley, 1996). The received *Máoshī* recension supplies the complete textual sequence used as the reference witness in this dataset (Máo Shī Zhèngyì, 2000). The poems have also played a central role in the study of early Chinese literature, ritual, language, and historical phonology.

Excavated manuscripts make it possible to compare the received text with material witnesses separated by date, provenance, and textual practice. The fourth-century BCE Āndà manuscript preserves 58 odes¹ and is the earliest substantial manuscript witness currently available (Huáng and Xú, 2019). Its textual stability and variation have already prompted renewed work on the formation, memorisation, and written transmission of the Odes (Smith & Poli, 2021; Meyer & Schwartz, 2022). The Fùyáng manuscript, placed in a tomb closed in 165 BCE, preserves fragmentary material from the early imperial period (Hú and Hán, 1988). Together, the received, Āndà, and Fùyáng witnesses provide a controlled body of parallel text spanning several centuries.

Existing editions and studies present transcriptions, commentary, and photographic plates, but they do not provide a shared machine-readable coordinate system in which corresponding occurrences can be compared directly. The present dataset supplies that layer.

The dataset is extensible to further witnesses as they become available in forms suitable for scholarly reuse. For a complete overview of excavated manuscripts preserving substantial material from the *Odes*, see Kern (2026, pp. 10–11). Shaughnessy (2025, p. xxviii) reports that the Xiàjiātái 夏家臺 manuscript contains 117 slips covering the first 14 poems of the *Airs of Bei*, while approximately 300 reconstructed Wángjiāzuǐ 王家嘴 slips contain passages from 154 poems. Preliminary publication also confirms that the Hǎihūnhóu 海昏侯 tomb contained an *Odes* manuscript (Zhū and Kē, 2021), and its variant readings have already supported phonological analysis (Yè, 2025), but the publication does not contain the manuscript images or a transcription that can be incorporated here. When that material, or transcriptions of other *Odes* manuscripts, becomes available, we will enter it in the same coordinate and alignment model as that of the present article.

2 METHOD

The dataset consists of three layers: the transcription of the witnesses, their physical locations in source editions, and finally correspondences across the witness transcriptions. The following sections explain how those layers are established and connected.

WITNESS SELECTION AND REFERENCE TRANSCRIPTIONS

The corpus currently contains three witnesses. The received *Máoshī* is based on the text published in the *Máoshī zhèngyì* 毛詩正義 (2000). The Āndà transcription uses Loh's (2022) study as its reference transcription and the published edition edited by Huáng and Xú (2019) as its physical source. The Fùyáng transcription uses Chén's (2019) collation as its reference transcription and the edition edited by Hú and Hán (1988) as its physical source.² The dataset's transcription files follow these upstream editions as closely as possible. Table 1 summarises the physical sources and reference transcriptions used for the three witnesses.

¹ We count the unaligned passage on the slip numbered 八十三, headed *Zuò yú zhī shī* 作魚之詩, as a poem, following Loh (2022, p. 279). It has no counterpart in either the received *Máoshī* or the Fùyáng witness, but future manuscript witnesses may preserve a parallel to this otherwise unaligned poem. This yields 58 poems, rather than the 57 odes preserved in whole or in part reported by Smith & Poli (2021, p. 530).

² Additional *Shijing* strips are reported in Chū (2005, vol. 14), and colour photographs of two strips appear in Fùyáng Shī Bówùguǎn (2022, plate 39). The available scans and photographs of the additional material were not adequate for image segmentation, although readings reported in the later scholarship are retained in Chén's (2019) transcription.

WITNESS	PERIOD	UNITS	PHYSICAL SOURCE	REFERENCE TRANSCRIPTION
Máoshī	Received	305	Received text	Máo Shī Zhèngyì (2000)
Āndà	Fourth century BCE	58	Huáng and Xú (2019)	Loh (2022)
Fùyáng	Early Western Han	113	Hú and Hán (1988)	Chén (2019)

Table 1 Witnesses and source works used in the corpus.

TRANSCRIPTION MODEL

Each witness’s transcription is stored in a versioned JSON file divided into ordered units (i.e., odes), paragraphs, and lines. We add punctuation and verse separation for display purposes as an editorial process. These do not reflect the punctuation of the manuscript, and they do not participate in textual alignment across witnesses.³ We encode the transcriptions using Unicode characters whenever a suitable character exists. A graph without a suitable encoded character is represented as one braced expression containing an Ideographic Description Sequence (IDS), the Unicode notation for describing a graph through the spatial arrangement of its components (Ideographic Description Sequences, 2026). For example, the third character extracted from strip 1 of the Āndà version of the *Guānjū* 關雎 (Huáng and Xú, 2019, p. 5) is shown in Figure 1. Loh (2022, p. 10) transcribes it as 駘; because this graph has no suitable Unicode code point, our dataset represents it as {𠂇鳥𠂇}: the character 𠂇 specifies a left–right arrangement of the components 鳥 and 𠂇. The braced IDS expression encodes one graph position. Damaged graphs can retain the same structural notation: for example, {𠂇 □ 區} in the Fùyáng witness of *Jìng nǚ* 靜女 records an unreadable left component beside 區; the generated font renders it as 區, preserving the surviving component and outlining the damaged region with a dashed border. We encode heavily damaged graphs as {?}, rendered as 區.⁴ Full documentation of IDS notation is outside the scope of this article. We transcribe iteration marks using Unicode’s OLD CHINESE ITERATION MARK (𠂇 U + 16FE3),⁵ and lacunae using WHITE SQUARE (□ U + 25A1) for missing single characters and MASU MARK (𠂇 U + 303C) for lacunae of unknown length. Figure 2 shows the structure of the transcription JSON file along with the first two lines of *Guānjū* 關雎 as found in the Āndà manuscript.



Figure 1 The third character Āndà’s *Guānjū* 關雎 Ode: 駘.

³ The transcriptions of Loh (2022) and Chén (2019) which we follow do not transcribe the actual manuscript’s punctuation and we have followed them for the present publication. This is an issue we will aim to address in a later update, so that textual segmentation and understanding of the text in these various traditions can be better studied.

⁴ The release supplies paired WOFF2 and TTF fonts generated by *ids-to-font* (Baley, 2026b), which derives its outlines from the Zi.tools API (Zi.tools, 2026); the archived fonts use commit 366cf6d.

⁵ See Richter (2023) for a discussion of all the usages of this mark.

```
{
  "schema_version": "1.0",
  "witness_id": "anda",
  "reference_transcription": "loh_chen-yee_2022_AnhuiDaxueCang",
  "title": "安大簡詩",
  "units": [
    {
      "type": "poem",
      "section": "周南",
      "title": "關雎",
      "paragraphs": [
        [
          "關 = 雎{鳥} ",
          "才河之州。",
          "..."
        ]
      ]
    }
  ],
  "..."
}
```

Figure 2 Opening extract of the Āndà transcription.json file. The ordered unit, paragraph, and line structure supplies stable coordinates for individual graph positions; ellipses abbreviate later lines and units.

IMAGE SEGMENTATION

The processing of manuscript images proceeds in two steps. First, for each strip, `strip_mapping.json` records the source-edition PDF page number and the strip's crop box, which consists of two pairs of (x,y) coordinates for the top-left and bottom-right corners of the rectangle that contains the image of the strip. This crop box is used to extract the image of the strip from the PDF. Second, `character_mapping.json` records the crop box, classification, and generated filename for each item — graph, section marker, etc. — on that strip. **Figure 3** illustrates these operations for the first Āndà strip of *Guānjū* 關雎. This layer describes image derivation from the PDF; it is independent of the transcription of the graph an item represents. The mappings retain the source coordinates alongside generated filenames, so strip and graph images can be rebuilt from a compatible source PDF when auditing or correcting crop geometry.

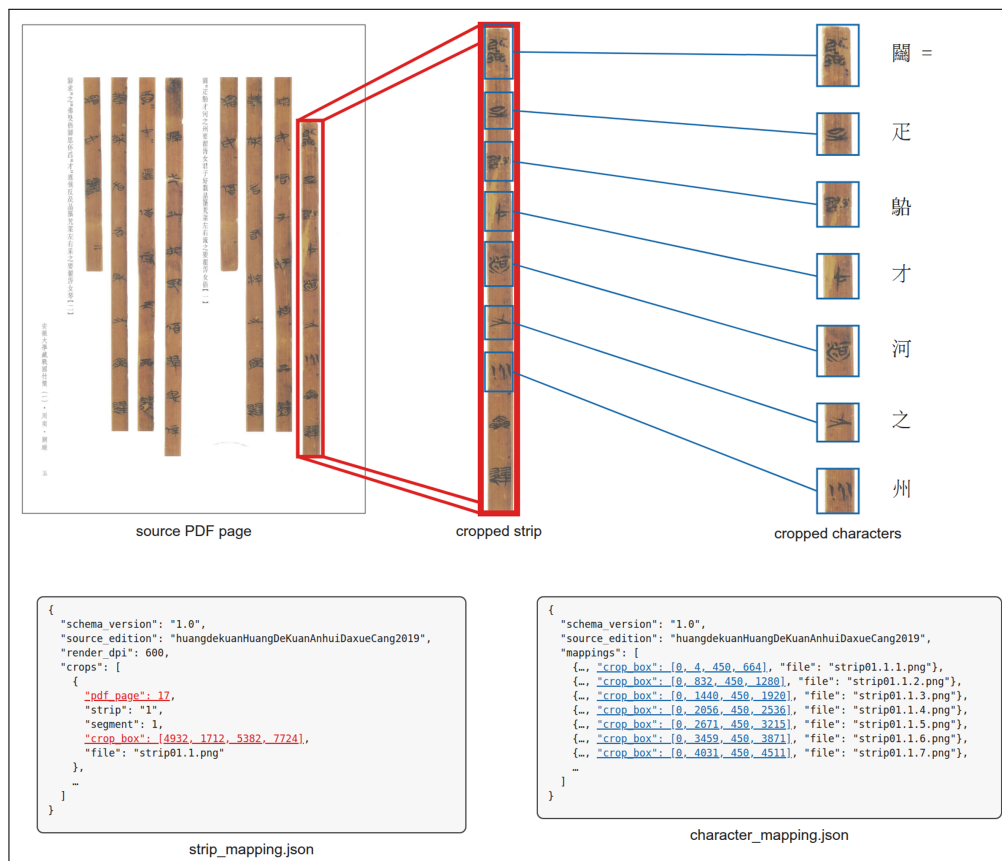


Figure 3 Image segmentation for the first strip of Āndà *Guānjū* 關雎. Red traces the source page selection to the reconstructed strip; blue traces selected strip locations to individual graph crops. The two mapping files preserve the geometry required to repeat both steps.

The classification of items, which distinguishes written graphs from strip numbers and section markers, provides useful physical metadata and could be extended to other forms of punctuation in the future.

The reviewed strip and graph crops are distributed with permission of the respective rights holders of the Āndà and Fùyáng manuscript images; the source PDFs and uncropped plates are not redistributed. Consequently, users can inspect the image layer without obtaining the source editions. The crop recipes remain available as an optional audit path: researchers with lawful access to a compatible source PDF can rebuild the images from the recorded geometry. The reconstruction checks the supplied PDF and resulting images against the reviewed media manifest, so a discovered crop error can be documented as a revised mapping and independently regenerated.

TRANSCRIPTION-TO-SOURCE MAPPING

`source_mapping.json` supplies the editorial connection between a manuscript's images and its transcription. Each mapping joins a transcription position to a physical occurrence by PDF page, strip, and item number. These mappings were created by hand in a project-specific curation tool, which is not part of this release. The release contains 2,876 Āndà transcription-to-source mappings and 3,004 source-to-image mappings. Fùyáng contains 888 transcription-to-source mappings and 897 source-to-image mappings. The difference between those two numbers is mainly due to strip numbers and other marks being accounted for in source-to-image mappings but not being transcribed. Additionally, some Fùyáng readings transcribed in Chén (2019) are known only through editorial reports of writings on the back of the strips or through later identified strips for which no usable plate image was available. These readings are retained without a source or image mapping. Figure 4 traces this relation for the opening of Āndà *Guānjū* 關雎: each cropped graph is connected, through its released source-mapping record, to its position in the rendered transcription.

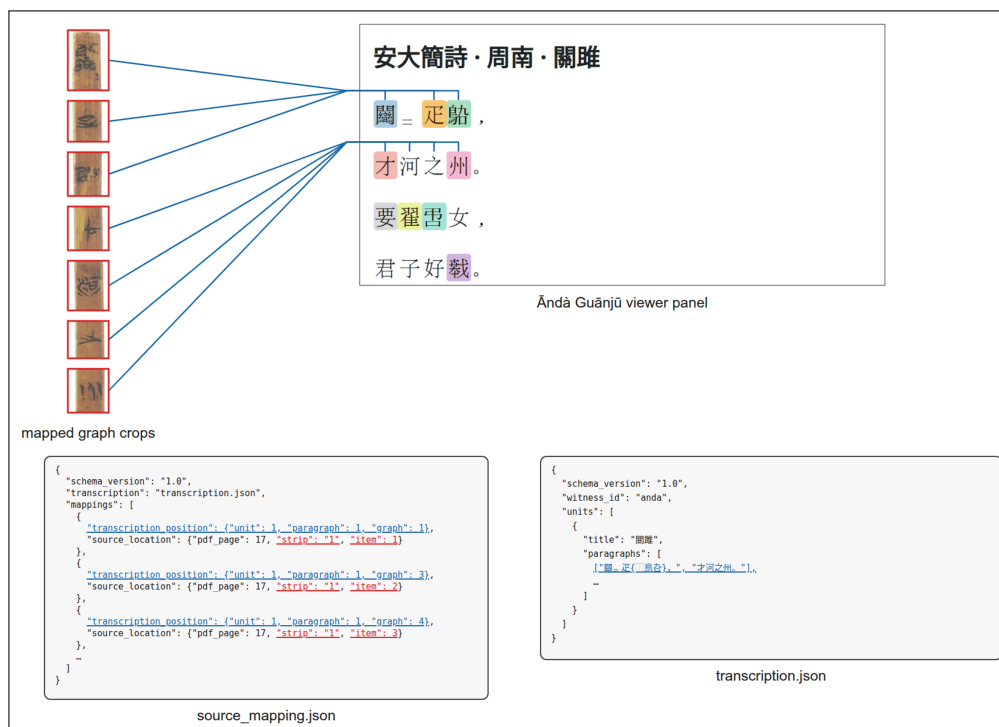


Figure 4 From physical graph crops to the Āndà *Guānjū* 關雎 transcription. Each source-mapping record connects a cropped graph to one position in the transcription: red identifies its physical location and blue its textual position. Colours are explained below in paragraph Visual inspection.

TRANSCRIPTION ALIGNMENT ACROSS WITNESSES

The separate `alignment.json` file records cross-witness transcription correspondence. Each alignment group contains two or more witness locations, each expressed as unit-paragraph-graph coordinates. This supports one-to-one, one-to-many, and many-to-many relationships between the witnesses. A preliminary alignment baseline was produced with GPT-5.6 Terra, then manually corrected and proofread by both authors of the present article. The

current release contains 3,466 correspondence groups comprising 7,319 graph locations. Of these groups, 3,256 connect two witnesses and 210 connect all three. **Figure 5** shows three three-witness groups for a section of the *Chún zhī bēnbēn* 鶉之奔奔 Ode in `alignment.json`, and the resulting view in our local web UI.

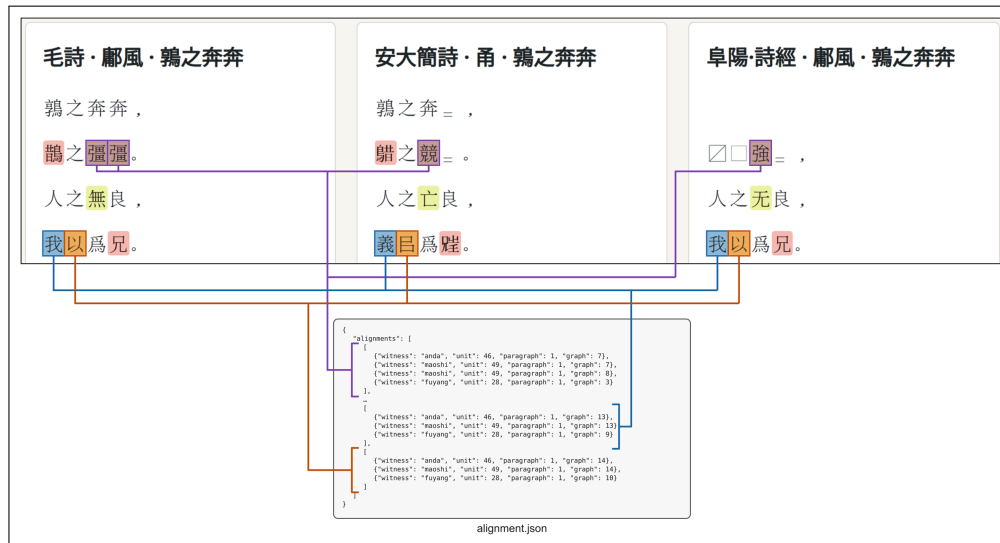


Figure 5 Cross-witness alignment in *Chún zhī bēnbēn* 鶉之奔奔. The purple correspondence relates Āndà競 and Fùyáng 強 to the two graphs 彊彊 in the received text. The blue and orange correspondences identify the first two graph positions of the received reading 我以爲兄 and its manuscript variants. Each array records the graph positions treated as corresponding. Manuscript images are hidden in this figure for legibility reasons.

EXTENSIBLE AND COMPETING TRANSCRIPTIONS

Witness identifiers are not restricted in principle to the three documents in the current release. A later publication could be added as another transcription file and linked to existing or new correspondence groups without altering the unit–paragraph–graph coordinate structure. The same mechanism can preserve more than one scholarly transcription of a manuscript: each transcription is represented as a separate witness, while its graph positions can map to the same physical strip and image locations. Competing graph identifications therefore remain independently citable and directly comparable. Independent studies of the Āndà manuscript already demonstrate how readings of particular graphs and passages can differ after renewed examination (Smith & Poli, 2021).

RELEASE STRUCTURE

The preceding layers are distributed as separate JSON files. Every witness, including the received Máoshī, has a `transcription.json` file. The two manuscript witnesses additionally have `source_mapping.json`, `strip_mapping.json`, and `character_mapping.json`; the first records editorial assignments, whereas the latter two record the page-to-strip and strip-to-graph crop operations illustrated in **Figure 3**. At the cross-witness level, `alignment.json` joins graph positions from the witness transcriptions into correspondence groups (**Figure 5**). The received Máoshī has only `transcription.json`, because it has no manuscript-image layer.

One last file, `expected_media.json`, acts as a reproducibility safeguard. It records the source-PDF digest and the expected identity of generated strip and graph images, allowing a media build to be checked without distributing the source PDF.

VALIDATION AND REPRODUCIBILITY

JSON Schemas define the transcription, alignment, source-mapping, image-mapping, and media-reconstruction formats. The validation suite that comes with the dataset checks schema conformance, bibliography references, graph coordinates, alignment uniqueness, mapping references, and media inventories. The release includes the reviewed strip and character images. Crop recipes can optionally reconstruct them from user-supplied source PDFs, while the expected-media manifests record file and decoded-pixel hashes, dimensions, and image modes. The release also includes a static viewer for poem-level browsing, graph selection, and graph search.

VISUAL INSPECTION

Along with the dataset, we provide a local browser interface for the visualisation of the data. The interface has two main views: first, its alignment view presents the received and manuscript witnesses in parallel, with a searchable poem selector. Navigation between poems can also be performed using left and right arrow keys. The alignment view highlights the correspondence between graphs across witnesses by assigning identical colours to them, and a toggle can filter this highlighting down to spelling variants only (i.e., identical graphs across witnesses are not highlighted). Selecting a graph or hovering over it highlights its linked locations. The graph image crops are shown alongside the transcription of each witness; clicking on any such graph image brings the user to the second view, the strip viewer.

In the strip view, shown in [Figure 6](#), the user is presented with a fully annotated strip: its graphs, section markers, and strip numbers are shown against the view of the full strip, so that the materiality of the strip can be better studied. Navigation between strips is supported through left and right arrow keys, and clicking on a graph takes the user to the corresponding alignment view. The interface therefore supports direct visual inspection while remaining a view onto the same JSON records available for computational reuse. This joins the manual and computational forms of exploration advocated for digital editions of bamboo manuscripts ([Pulini and List, 2025](#)).

Finally, a basic corpus-wide graph search feature lets the user search for encoded graphs, braced IDS expressions, and the iteration mark, and links results back to their poem alignments.

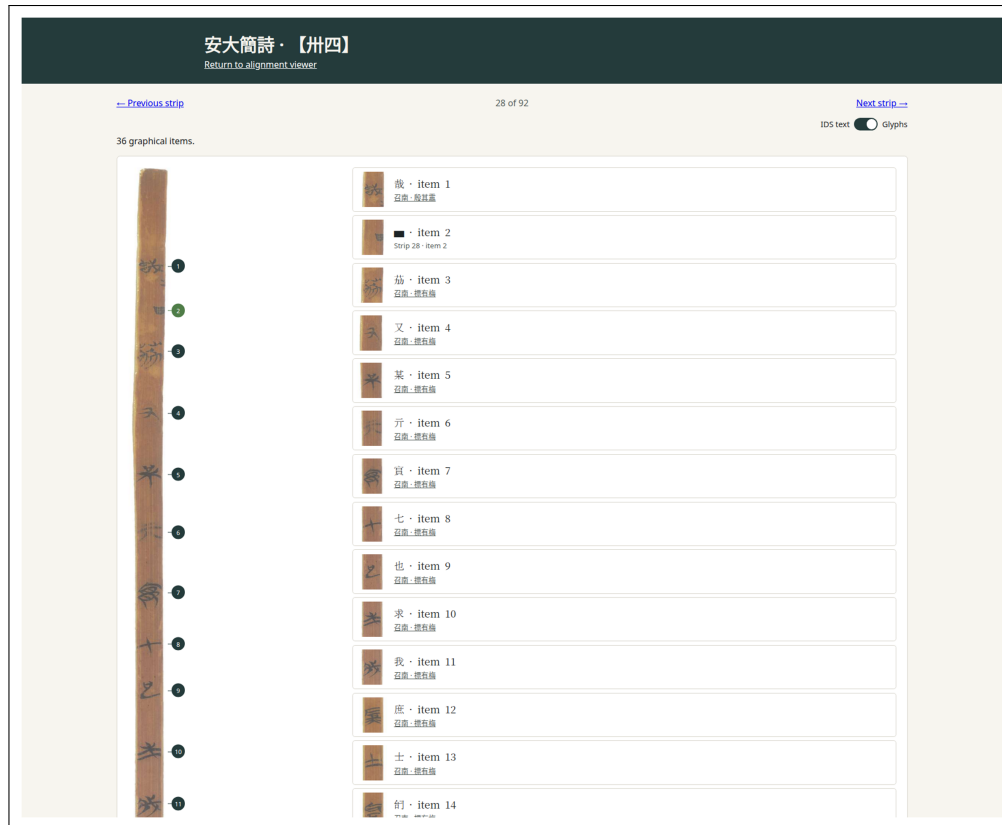


Figure 6 Āndà strip view with physical item overlays and crop metadata.

3 DATASET DESCRIPTION

OBJECT NAME

Aligning the Odes: A Character-Level Parallel Dataset for the Anhui University and Fùyáng Bamboo-Slip Manuscripts.

FORMAT NAMES AND VERSIONS

UTF-8 JSON, JSON Schema Draft 2020-12, HTML, CSS, JavaScript, Markdown, PNG, WOFF2, TrueType, BibTeX, and Citation File Format.

CREATION DATES

2026-08-06 to 2026-08-29.

Baley and Ying
*Journal of Open
Humanities Data*
DOI: 10.5334/johd.644

8

DATASET CREATORS

Julien Baley, SOAS University of London: Conceptualization, Data curation, Methodology, Software, Validation, and Visualization. Ying Jinqí 應金琦, Trinity College Dublin: Data curation, Investigation, and Validation.

LANGUAGES

Classical Chinese and English documentation.

LICENCE

Creative Commons Attribution 4.0 International applies to project-authored data, code, schemas, documentation, and reproducibility recipes. Third-party source PDFs and uncropped manuscript plates are excluded. Reviewed strip and graph crops are distributed under permission from their respective rights holders and are not relicensed under CC BY 4.0. Permission to publish the project transcriptions based on Chén (2019) and Loh (2022) was obtained from their respective authors. Permission to publish the Āndà crops derived from Huáng and Xú (2019) was obtained from Huáng Dékuān 黃德寬 and the Chinese Character Development and Application Research Center (Ānhuī Dàxué Hànzì Fāzhǎn yǔ Yìngyòng Yánjiū Zhōngxīn 安徽大學漢字發展與應用研究中心). Permission to publish the Fùyáng crops derived from Hú and Hán (1988) was obtained from Shànghǎi gǔjí chūbǎnshè 上海古籍出版社. The release bibliography and image-rights statement record their provenance and reuse terms.

REPOSITORY NAME

Zenodo.

PUBLICATION DATE

26 September 2026.

DATASET INVENTORY

The release contains complete transcriptions of 305 Máoshī poems, 58 Āndà poems, and 113 Fùyáng units: 76 identified poems and 37 entries in the separately identified fragment appendix. These transcriptions contain 29,646 Máoshī, 3,185 Āndà, and 1,416 Fùyáng graph positions. The alignment file links 2,805 Āndà, 884 Fùyáng, and 3,630 Máoshī graph locations.

CONTRIBUTING

Please email corrections and proposed improvements to the corresponding author.

4 RE-USE POTENTIAL

The dataset supports direct comparison of corresponding passages across the three major *Shijing* witnesses. Researchers can retrieve all recorded variants for a graph or poem, distinguish graphical variation from additions and omissions, and trace each mapped manuscript reading back to its physical source location. Because correspondence is stored independently of character identity, the same data can support philological review without treating variant graphs as automatically equivalent or different. The viewer makes that review accessible without specialist software: a proposed reading can be checked against its corresponding passage and, where available, its reconstructed manuscript image. This combination of structured annotation and direct inspection responds to the complementary manual and computational uses of bamboo-slip data identified by Pulini and List (2025): it provides a structured foundation for full digital critical editions, on which projects can add critical apparatus, discussions of readings, translations, and glosses.

For textual history and manuscript studies, the corpus makes the distribution of stable and unstable passages measurable across witnesses. For palaeography, IDS expressions retain analysable structure where conventional Unicode transcription is insufficient. The recorded crop geometry also lets a researcher inspect a disputed crop, propose a corrected mapping, and, with lawful access to the cited source edition, independently regenerate the image. The reviewed image–transcription links also provide a controlled corpus for testing graph retrieval and handwritten-character recognition methods while keeping outputs traceable to the underlying documentary evidence.

For historical phonology, the aligned positions provide a controlled inventory of spelling variants. These variants can act as evidence for sound mergers or distinction maintenance. A prototype version of this dataset informed the statistical evaluation of Warring States and Han rhyme evidence in Baley (2026a).

The viewer and structured files support complementary forms of reuse. A reader can inspect a poem, its differences, the available graph images, and their positions on reconstructed strips without specialist software. Researchers can use the same records programmatically for corpus queries, statistical analysis, alignment methods, and variant-graph retrieval. The graph-level correspondences also provide a fine-grained basis for investigating structural variation between witnesses, including differing segmentation, expansion or contraction, and stanzaic organisation.

The corpus can grow without replacing earlier releases. We commit to integrating the known Odes manuscripts from Hǎihūnhóu, Xiàjiātái, and Wángjiāzuǐ when their images and transcriptions are published in forms suitable for scholarly reuse. Alternative transcriptions of a witness can coexist as separate documents linked to the same source locations and images. Users can then compare readings with one another and with the underlying graphical evidence. The source and image layers also allow projects with different media rights to share compatible textual and provenance data.

The framework is not restricted to the Odes. We also intend to adapt its separation of transcription, physical evidence, and editorial correspondence to other early Chinese textual traditions, beginning with the *Lǎozǐ* 老子 and the *Yìjīng* 易經 (*Book of Changes*). Such applications would make it possible to compare related manuscript and received-text witnesses without imposing a single normalised transcription on them. For the *Yìjīng*, we intend to align the received *Zhōuyì* 周易 with the Mǎwángduī 馬王堆, Fùyáng, Shanghai Museum, and Hǎihūnhóu witnesses, and to include related divination manuals, such as the Wángjiātái *Guīcáng* 歸藏 (Shaughnessy, 2014), the Shanghai Museum *Bǔ shū* 卜書 (Mǎ, 2012), or the Tsinghua *Shì fǎ* 筮法 and *Bié guà* 別卦 (Lǐ, 2013). This broader comparison will make it possible to study parallel passages across divination manuals as well as variation within the *Zhōuyì* tradition.

5 LIMITATIONS

The alignment represents the present editors' correspondence judgements and does not eliminate interpretive uncertainty. Damaged text, iteration marks, and one-to-many relations require case-by-case philological interpretation. The received witness is much larger than either manuscript witness, so most Máoshī graph positions necessarily remain unlinked. Fùyáng's fragment appendix is transcribed but not yet aligned or source-mapped, and a small number of readable Fùyáng positions lack identified image evidence. The release excludes third-party source PDFs and uncropped plates; users who independently reconstruct the permissioned crop set must obtain the cited editions. The current media derives from Hú and Hán (1988); more recent photographs of some Fùyáng strips are published in Chū (2005) and Fùyáng Shì Bówùguǎn (2022), but the available reproductions did not support systematic image segmentation for this release. The current viewer is configured for the three released witnesses; displaying later witnesses or competing transcriptions will require extending its witness configuration.

ACKNOWLEDGEMENTS

The authors would like to thank the following: Chén Chén 陳晨 and Loh Chen-Yee 駱珍妮 for permission to use their transcriptions, Professor Huáng Dékuān 黃德寬 and the Chinese

Character Development and Application Research Center (Ānhuī Dàxué Hànzì Fāzhǎn yǔ Yīngyòng Yánjiū Zhōngxīn 安徽大學漢字發展與應用研究中心) for permission to include the images of the Āndà manuscript, and Shànghǎi gǔjí chūbǎnshè 上海古籍出版社 for permission to include the images of the Fùyáng manuscript.

FUNDING STATEMENT

This work was funded by the Arts and Humanities Research Council (AHRC), UKRI, as part of the project “Han Phonology: When Chinese Became Chinese.” Project Reference: AH/V008722/1. Principal Investigator: Nathan Hill, SOAS University of London and Trinity College Dublin.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Julien Baley: Conceptualization, Data curation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. Ying Jinqi 應金琦: Data curation, Investigation, Validation, Writing – original draft, Writing – review & editing.

AUTHOR AFFILIATIONS

Julien Baley  orcid.org/0000-0003-1056-6211
School of Languages, Cultures, and Linguistics, SOAS University of London, UK

Ying Jinqi 應金琦  orcid.org/0000-0002-6154-0620
Trinity Centre for Asian Studies, Trinity College Dublin, Ireland

REFERENCES

- Baley, J. (2026a, July 22). Did Old Chinese *-r Exist? A statistical re-evaluation of Warring States and Han rhyme evidence. *Conference presentation, 26th Biennial Conference of the European Association for Chinese Studies*. Venice.
- Baley, J. (2026b). *ids-to-font: Build IDS ligature fonts and encoded Unicode supplements from Zi.tools outlines*. Software, version 0.1.0. Retrieved from <https://github.com/julienbaley/ids-to-font>
- Chinese Character Description Languages: Ideographic Description Sequences*. (2026). Wikipedia, The Free Encyclopedia. Retrieved from https://en.wikipedia.org/wiki/Chinese_character_description_languages#Ideographic_Description_Sequences. (Accessed 12 August 2026).
- Chén Chén 陳晨. (2019). *Jiǎn bó “Shī” jí xiāngguān wénxiàn zhěnglǐ yǔ yánjiū* 簡帛《詩》及相關文獻整理與研究. (Unpublished doctoral dissertation). Wuhan University, Wuhan.
- Chū Shibīn 初師賓 (Ed.) (2005). *Zhōngguó jiǎndú jíchéng* 中國簡牘集成, (Vol. 14). Lanzhou: Dūnhuáng Wényì Chūbǎnshè 敦煌文藝出版社.
- Fùyáng Shì Bówùguǎn 阜陽市博物館 (2022). *Fùyáng Shuānggǔduī Hànmù* 阜陽雙古堆漢墓. Beijing: Zhōnghuá Shūjú 中華書局.
- Hú Píngshēng 胡平生, & Hán Zìqiáng 韓自強 (Eds.) (1988). *Fùyáng Hànjǎn Shījīng yánjiū* 阜陽漢簡詩經研究. Shanghai: Shànghǎi Gǔjí Chūbǎnshè 上海古籍出版社.
- Huáng Dékuān 黃德寬, & Xú Zàiguó 徐在國 (Eds.) (2019). *Ānhuī Dàxué cáng Zhànguó zhújiǎn (1)* 安徽大學藏戰國竹簡 (一) (*Warring States Bamboo Slips at Anhui University – 1*). 中西書局 (Zhongxi Book Company).
- Kern, M. (2026). *Writing and Community in 300 BCE China – The Canonical Poems (Shi 詩) at the Chu 楚 Capital of Ying 郢*. Leiden: Brill. <https://doi.org/10.1163/9789004759022>
- Lǐ Xuéqín 李學勤 (Ed.) (2013). *Qīnghuá Dàxué cáng zhànguó zhújiǎn (sì)* 清華大學藏戰國竹簡 (肆) (*Warring States Period Bamboo Slips from Tsinghua University Collection – Volume 4*). Zhōngxī shūjú 中西書局.
- Loh Chen-Yee 駱珍伊 (2022). *Ānhuī Dàxué cáng Zhànguó zhújiǎn “Shījīng” yánjiū* 安徽大學藏戰國竹簡《詩經》研究. (Unpublished doctoral dissertation). National Taiwan University, Taipei.
- Mǎ Chéngyuán 馬承源 (Ed.). (2012). *Shànghǎi Bówùguǎn cáng zhànguó Chǔ zhú shū 9* 上海博物館藏戰國楚竹書 (九) (*The Warring States Chu Bamboo Slips in the Shanghai Museum – 9*). Shànghǎi gǔjí chūbǎn shè 上海古籍出版社.
- Máo Shī zhèngyì 毛詩正義. (2000). Beijing: Běijīng dàxué chūbǎn shè 北京大學出版社.

- Meyer, D., & Schwartz, A. C. (2022). *Songs of the Royal Zhōu and the Royal Shào: Shī 詩 of the Ānhuī University Manuscripts*. Leiden: Brill. <https://doi.org/10.1163/9789004512436>
- Pulini, M., & List, J.-M. (2025). Using Cross-Linguistic Data Formats to Enhance the Annotation of Ancient Chinese Documents Written on Bamboo Slips. In *Proceedings of the Second Workshop on Ancient Language Processing* (pp. 31–37). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.alp-1.4>
- Richter, M. L. (2023). Towards a Broad Concept of Punctuation. *Bamboo and Silk*, 6(2), 159–185. <https://doi.org/10.1163/24689246-20230032>
- Shaughnessy, E. L. (2014). *Unearthing the Changes: Recently Discovered Manuscripts of the Yi Jing (I Ching) and Related Texts*. New York: Columbia University Press. <https://doi.org/10.7312/shau16184>
- Shaughnessy, E. L. (2025). *The Classic of Poetry: Ancient China's Songbook*. Hong Kong: Chinese University of Hong Kong Press.
- Smith, A., & Poli, M. (2021). Establishing the text of the Odes: The Anhui University Bamboo Manuscript. *Bulletin of the School of Oriental and African Studies*, 84(3), 515–557. <https://doi.org/10.1017/S0041977X22000015>
- Waley, A. (1996). *The Book of Songs: The Ancient Chinese Classic of Poetry*. Grove Press.
- Yè Yùyīng 葉玉英 (2025). Cóng Hǎihūnhóu mù Shījīng yìwén kàn Xihàn zhōngqī Hànyǔ de ruògān shēngyùn tèzhēng 從海昏侯墓《詩經》異文看西漢中期漢語的若干聲韻特徵. *Conference presentation, Recent Advances in Chinese Historical Phonology*. University of Vienna. Retrieved from <https://www.youtube.com/watch?v=JdCszehtdo>
- Zhū Fènghàn 朱鳳瀚, & Kē Zhōnghuá 柯中華 (Eds.) (2021). *Hǎihūn jiǎndú chūlùn* 海昏簡牘初論. Beijing: Běijīng Dàxué Chūbǎnshè 北京大學出版社.
- Zi.tools 字統網. (2026). Web application and API. Retrieved from <https://zi.tools/>. (Accessed 12 August 2026).

TO CITE THIS ARTICLE:
Baley, J., & Ying, J. 應金琦. (2026). Aligning the Odes: A Character-Level Parallel Dataset for the Anhui University and Fùyáng Bamboo-Slip Manuscripts. *Journal of Open Humanities Data*, 12: 138. DOI: <https://doi.org/10.5334/johd.644>

Submitted: 28 August 2026
Accepted: 23 September 2026
Published: 19 October 2026

COPYRIGHT:
© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <https://creativecommons.org/licenses/by/4.0/>.

Journal of Open Humanities Data is a peer-reviewed open access journal published by Ubiquity Press.