Biocatalytic routes to stereo-divergent iridoids

Thousands of natural products are derived from the fused cyclopentane-pyran molecular scaffold nepetalactol. These natural products are used in an enormous range of applications that span the agricultural and medical industries. For example, nepetalactone, the oxidized derivative of nepetalactol, is known for its cat attractant properties as well as potential as an insect repellent. Most of these naturally occurring nepetalactol-derived compounds arise from only two out of the eight possible stereoisomers, 7S-cis-trans and 7R-cis-cis nepetalactols. Here we use a combination of naturally occurring and engineered enzymes to produce seven of the eight possible nepetalactol or nepetalactone stereoisomers. These enzymes open the possibilities for biocatalytic production of a broader range of iridoids, providing a versatile system for the diversification of this important natural product scaffold.

Iridoids are a class of monoterpene natural products that are produced in a diverse range of flowering plant species. Over one thousand distinct iridoid structures have been identified, and collectively, these compounds play wide-ranging roles in plant defense. Moreover, many iridoids have pharmaceutical applications, including iridoid-derived monoterpene indole alkaloids such as the well-known anticancer drug vincristine. A key intermediate in the generation of most iridoids is nepetalactol, which, along with its oxidized product, nepetalactone, is well-known for its behavioral effect on cats. The nepetalactol stereoisomers found in nature are 7S-cis-cis, 7R-cis-cis, 7S-cis-trans, 7R-cis-trans, 7S-trans-cis, 7R-trans-cis, and 7S-trans-trans. However, plants in the genus Nepeta produce iridoids with varied stereochemistry at the bridgehead carbons; in addition to 3a, iridoids with 7S-cis-cis (7S, 4aS, 7aS) 3b and 7S-trans-cis (7S, 4aS, 7aS) 3c stereochemistry are observed. We recently discovered a family of enzymes in Nepeta mussinii and Nepeta cataria, nepetalactol-related short chain reductases or NEPS, that, in tandem with a 7S-specific ISY, control the oxidation of GPP to form the 7S-cis trans nepetalactol scaffold, the precursor to all iridoids in plants. The 7S-cis-trans nepetalactol scaffold has three chiral centers, though only a subset of the 8 possible stereoisomers have been identified in nature.

As with the vast majority of monoterpene pathways, iridoid biosynthesis begins with geranyl pyrophosphate (GPP), though in iridoid biosynthesis, GPP is invariably converted to geraniol, and then oxidized to 8-oxogeranial. The ISY product, to form the 7S-cis-trans nepetalactol scaffold, the precursor to all iridoids in plants.
NEPS belong to the NAD+-dependent short chain dehydrogenase family (SDR) and share 65–95% amino acid identity among them as veol dehydrogenase. Some NEPS, although containing an NAD+ as well as 62–73% identity to was selected since it is the most evolutionarily distant member of the Nature Communications | (2022) 13:4718 point for developing stereo-selective cyclases to enable this. catalytic and synthetic systems. The NEPS provide a starting point for developing stereo-selective cyclases to enable this.

Here, we set out to understand and expand the selectivity of NEPS by both gene discovery in Nepeta sibirica and enzyme engineering of NEPS from various Nepeta species: N. sibirica, N. cataria and N. mussinii. To compile a family of NEPS with the greatest possible natural catalytic variety, we identify six additional NEPS homologs in Nepeta sibirica, N. mussinii, N. cataria, and N. mus- 

Fig. 1 | Overview of the nepetalactol scaffold and its stereo-control in the Nepeta species. A. The early iridoid pathway takes two main directions based on the preferred stereocchemistry of iridoid synthase (ISY). 7S-ISY leads to biosynthesis products such as (7S) nepetalactones found in most Nepeta species, as well as well-known indole alkaloid precursor strictosidine in C. roseus and other species (red). 7R-ISY exists in other members of the Asterids clade, particularly the Lamiales order, where iridoid glucosides with this C7 stereocchemistry is common (light blue). B. Nepetalactol 3 stereocchemistry can adopt 8 stereo-chemical configurations based on the relative positions of the C7 methyl group and the 7a-4a bridge. Most iridoids known in plants are based on the 7S-cis-trans 3a and 7R-cis-cis 3b’ configuration (7S, 4aS, 7aR and 7R, 4aS, 7aR, respectively). C. Three Nepeta species studied here, N. sibirica, N. mussinii, and N. cataria. D. As for nepetalactol 3, nepetalactone 4 also can exist in 8 stereo-chemical configurations. NEPS typically control the stereocchemistry of the bridgehead carbons; in one instance a Major Latex Like Protein catalyzes the formation of 7S-cis-trans 3a (7S, 4aS, 7aR). NEPS also must be capable of cyclization and oxidation of both the 7S and 7R products of ISY in a selective manner to generate the full complement of stereoisomers.

Stereo-selective cyclization of 2 to generate these three nepetalactol isomers (Fig. 1D). In addition, the crystal structure of N. mussinii NnNEPS3, which catalyzes cyclization and partial oxidation to the 7S-cis-cis nepetalactone 4b isomer, was previously reported (PDB 6F9Q). NEPS belong to the NAD+ dependent short chain dehydrogenase family (SDR) and share 65–95% amino acid identity among them as well as 62–73% identity to Mentha x piperita (-) isoepipertol/(-) carvool dehydrogenase. Some NEPS, although containing an NAD+ binding site, are redox inactive and catalyze only cyclization, others catalyze only the oxidation from nepetalactol to nepetalactone, while others are dual-function enzymes, catalyzing both stereo-selective cyclization and oxidation of various nepetalactols to nepetalactones. Enzymatic formation of diverse iridoid stereoisomers could lead to the synthesis of natural and derivatives of bioactive compounds in biocatalytic and synthetic biology systems. The NEPS provide a starting point for developing stereo-selective cyclases to enable this.

Results and discussion
Cyclization and oxidative selectivity in NEPS
With a total of 19 wild-type NEPS sequences ranging from 64–96% sequence identity and two 3D NEPS structures to compare and contrast (Supplementary Figs. 1–2) we sought to understand the. We first set out to decouple oxidation and cyclization activity of NEPS through mutagenesis to maximize biocatalytic flexibility. In addition to cyclization of 8-oxocitronellyl enol 2 to nepetalactol 3, many NEPS catalyze the oxidation of 3 to nepetalactone 4 (Fig. 1D). We searched for

https://doi.org/10.1038/s41467-022-32414-w
mutations that would effectively decouple these cyclization and oxidative activities. Since NEPS belong to the NAD⁺-dependent short chain dehydrogenase family, mutation of the active site tyrosine residue required for oxidation in this class of enzymes could be a straightforward way to decouple cyclization and oxidative activity. Biochemical assays showed that NsNEPS2 (N. sibirica) is a cis-trans cyclase and oxidase (Fig. 2A); when NsNEPS2 is incubated with 8-oxogeranial 1 and 7S-isy, 7S-cis-trans nepetalactone 4a, with no presence of iridodial 5 byproducts, is observed. It is important to note that in the absence of a cyclase, 1 and 7S-isy yield the spontaneous cyclization product, 7S-cis-trans nepetalactol 3a. However, numerous byproducts, most notably partially cyclized iridodials 5, are observed in the spontaneous
reaction (Fig. 1D), and therefore a cis-trans cyclase can be identified by the absence of these byproducts in the reaction mixture. Mutation of the active site tyrosine in NsNEPS2 (NsNEPS2-Y167F) yielded an enzyme that produces only 7S-cis-trans nepetalactol 3a, with no irido-dial byproducts, suggesting that cyclization activity was maintained while oxidation activity was eliminated (Fig. 2A). However, this did not prove to be a general strategy for decoupling of cyclization and oxidation activity in other NEPS, since mutation of the catalytic tyrosine residue in other NEPS enzymes yielded enzymes that were completely inactive or, surprisingly, unchanged in oxidation activity in NmNEPS1 and NnNEPS1B (Supplementary Fig. 3). In NmNEPS1 and NsNEPS1B there is clearly a compensatory, albeit subtle, mechanism that allows oxidation even in the absence of this tyrosine. One possibility is that as long as the substrate is positioned appropriately next to the cofactor, another residue in the enzyme molecule could serve as the base that would be required to facilitate this oxidation of the alcohol. However, with limited structural information on the NEPS it is not possible to pinpoint these subtle differences. Therefore, we focused on exploring changes to the rest of the binding pocket while leaving the catalytic residues intact.

To search for alternative mutations to decouple cyclization and oxidation, NmNEPS3 (N. musáinii), which produces 7S-cis-cis nepetalactol 3b and only traces of the oxidized 7S-cis-cis nepetalactone 4b, was compared with NcNEPS3A (N. cataria 93% amino acid identity, Supplementary Fig. 1), which carries out both cyclization and oxidation (Fig. 1), and which carries out both cyclization and oxidation to 4b (Fig. 2B). This comparison suggested that residue 206 (Q in NmNEPS3, V in NcNEPS3A) could be responsible for the difference in oxidase activity. The mutant NmNEPS3-Q206V produces similar levels of 4b as NcNEPS3A, and NcNEPS3A-V206Q produces only 3b (Fig. 2B). Additional amino acid substitutions at residue 206 suggest that the polarity and size of the sidechain affect the oxidation of nepetalactol (Supplementary Fig. 4). Given the proximity of this residue to the binding site as evidenced by molecular docking (Fig. 2C), this residue may impact the exact orientation in which 3b binds, thereby determining whether the lactol group is positioned such that it can be oxidized by the NAD+ cofactor. Thus we can control dehydrogenase activity but not impact cyclization in the 7S-cis-cis cyclase NEPS3 through a single substitution at residue 206.

Sequence comparison of these highly similar 7S-cis-cis cyclase enzymes, NmNEPS3 and NcNEPS3A, with that of other NEPS also revealed the identification of a variable loop region in the binding pocket (Supplementary Fig. 5A, B). Mutation of this loop in NcNEPS3A (from N150 to V162) with the corresponding residues from NmNEPS3, an enzyme known to produce 7S-cis-trans nepetalactone 4a, resulted at successful swap of specificity from 4b to 4a (Supplementary Fig. 5C), showing that this loop plays a role in shaping the binding pocket and thereby setting the stereo-selectivity of cyclization. However, when the same residues from NmNEPS4, a 7S-trans-cis 3c cyclase, were introduced into NcNEPS3A, no 7S-trans-cis product was observed, albeit some 4a was observed. We concluded that while the ISO-162 loop is important in determining stereo-selectivity, it does not fully control it. Therefore, we needed to examine the role other regions of the binding pocket play in determining specificity.

To identify other regions responsible for controlling cyclization selectivity, we examined NmNEPS1 and NmNEPS4 due to the distinct and seemingly co-dependent catalytic activities (Fig. 3A) and high amino acid identity of these two enzymes (81% amino acid identity). NmNEPS4 cyclizes 8-oxocinornetyl enol 2 to form 7S-trans-cis nepetalactol 3c, which in the absence of a dehydrogenase spontaneously opens to form 7S-trans-cis iridodial 5c,d,e. NmNEPS1 has no cyclization activity, but NmNEPS1 is able to oxidize both 3c and 7S-trans-nepe-talactol 3a. NmNEPS4 can only oxidize 3a (Fig. 3A). Comparison of the binding pockets of these two enzymes revealed differences in eight residues with proximity to the binding pocket, six of which are located in the previously identified loop region in NEPS3 above (resides 152-164 in NmNEPS1 case, plus positions 1504 and 1598 in NmNEPS1 (equivalent to A105 and L199 in NmNEPS4) (Supplementary Fig. 6A, B). We replaced these 8 residues in NmNEPS1 (dehydrogenase activity only) with the residues found in NmNEPS4 (3c cyclase); gratifyingly, this resulted in a variant capable of generating 5c iridodial, successfully demonstrating that this cyclization activity can be installed into a dehydrogenase (Fig. 3A, Supplementary Fig. 6C). A variety of mutant combinations showed that 5 mutations, I545ATA (in the loop region), along with L198, is the minimum requirement for formation of 5c in the context of NmNEPS1. Based on the available NEPS crystal structures (NmNEPS3 and NsNEPS2), we hypothesized that the relatively bulky L199 residue in NmNEPS4 prevents the orientation of the 3c relative to the NAD cofactor required for oxidation (Fig. 3C). As we predicted, reversing mutation of L198 back to serine resulted in a NmNEPS1 variant capable of both cyclizing and oxidizing to 7S-trans-cis nepetalactone 4c, a catalytic activity not seen in the native NEPS. Extensive mutagenesis of the 1545ATA region (Supplementary Fig. 7) as well as the L198 residue (Supplementary Fig. 8) impacted the ratio of 4c and 4a product ratios. In particular, I545VTA mutant appears to improve the ratio towards 4c, minimizing the amount of spontaneously formed 4a (Fig. 3B). The I545ATA region enables the switch between cis-trans and trans-cis cyclization by shaping the binding pocket, though again, this switch must act in the context of two highly similar protein backgrounds.

Discovery of trans-trans cyclase in N. sibirica

The 7S-trans-trans nepetalactol 3d or nepetalactone 4d isomers were not observed in any of these mutagenesis experiments, prompting us to search for the naturally occurring cyclase that produces this stereoisomer. Based on achiral GC-MS analysis, we observed a compound that appeared to have 4d, in the phylogenetically distant N. sibirica, prompting us to search for cyclases with 7S-trans-trans stereo-selectivity in this plant. We identified and functionally characterized all ISY, NEPS, and MLPL orthologues in the N. sibirica leaves transcriptome (Supplementary Tables 2–3). First, orthologues of ISY and the closely related progesterone 5S-reductase (PSBR) from N. sibirica were assayed and analyzed in a GC chiral column, revealing that all of these enzymes reduced 8-oxogeranial 1 with 7S specificity at C7 (Fig. 4A), the same as ISY enzymes from C. roseus as well as N. cataria and N. mussinii. A relative ISY from the Lamiaceae species Lamium album (LalISY) was identified and used to generate the products with 7R stereochemistry. Thus, we initially tested the activity of all putative N. sibirica cyclases with the N. sibirica iridoid synthase in search of 7S-trans-trans nepetalactone activity.

As described above, we identified, characterized and subjected to crystallography analysis the NEPS orthologue NsNEPS2 as a 7S-cis-trans nepetalactone 4a synthase. Additional NEPS orthologues NsNEPSL,
NsNEPS1A, NsNEPS1B, NsNEPS4A, and NsNEPS4B were also identified (Fig. 4B, Supplementary Figs. 1–2). No orthologs of NEPS3 (7S-cis-cis) or NEPS5 (7S-cis-trans) were found. RNA-seq data suggests that NsNEPS1A, NsNEPS1B, NsNEPS4B are the most highly expressed genes, with NsNEPS1A being the highest (Fig. 4C).

Enzyme assays with 8-oxogeranial and NsISY show that NsNEPS1A, NsNEPS1B, NsNEPS2 and NsNEPSL all oxidize 7S-cis-trans nepetalactol 3a to 7S-cis-trans nepetalactone 4a (Fig. 4D, E). NEPS4 (NsNEPS4A and NsNEPS4B) lacked any activity under these conditions. NsMLPL1 appeared to be a 3a cyclase while the other MLPL homologs did not show activity towards 8-oxocitronellyl enol 2 (Fig. 4D, E). However, NsNEPS1B produced an unprecedented stereo-chemical profile, with all four of the 7S nepetalactone isomers, including 7S-trans-trans nepetalactone 4d, being observed in the reaction mixture (Fig. 4D). Curiously, although activity was achieved, the ratio in which it was produced was not consistent with the profile observed in N. sibirica. Upon reviewing the literature, we found a report demonstrating the existence of 7R-trans-trans nepetalactone 4d in N. elliptica\(^2\). Although NsISY is in fact 7S specific, we decided to rely on the 7R specific LaISY to test this hypothesis.
Reactions with 7R-ISY suggest the role of NEPS in *N. sibirica*

Given the prevalence of iridoids with 7R stereochemistry outside of the *Nepeta* genus, we tested the cyclization and oxidation activity of the NEPS in combination with 7R-ISY to determine whether NEPS could also catalyze the stereo-selective cyclization/oxidation of 7R iridoid scaffolds. 7R-ISY from the plant *Lamium album* (LaISY) was used as the 7R-ISY (Figs. 4A, 5A). NsNEPS1B, which can produce 7R-trans-trans nepetalactone, produced both 7R-trans-trans nepetalactone 4d' and 7R-cis-trans nepetalactone 4a' when assayed with LaISY and 8-oxogeranial 1 (Fig. 5B). The most highly expressed NEPS in *N. sibirica*, NsNEPS1A made exclusively 7S-cis-trans nepetalactone 4a when assayed with 7S-ISY (Fig. 4D), but in combination with LaISY, NsNEPS1A...
produced 7R-trans-trans nepetalactone 4d as the main product, along with lower amounts 7R-cis-trans nepetalactone 4a (Fig. 3B). Efforts to decouple cyclization and dehydrogenation of these enzymes via catalytic mutation (Supplementary Fig. 3) or via binding pocket redesign (Supplementary Fig. 9) did not yield enzymes with substantially different catalytic activities albeit changes in the ratio of nepetalactones were observed.

These enzymatic results led us to re-inspect the nepetalactone profile of *N. sibirica* using a combination of chiral GC column (Fig. 3B). Chemical epimerization of 7-cis-cis nepetalactone 4b derived from *N. mussinii* and putative *N. sibirica* trans-trans nepetalactone (Fig. 5C; Supplementary Figs. 10, and 14–17), and circular dichroism analysis of the resulting enantiomers (Fig. 5D). Careful analysis lead us to confirm that the leaves of *N. sibirica* contained 7R-trans-trans nepetalactone 4d and not 7S-trans-trans nepetalactone 4d, as we had initially assumed upon initial inspection of the leaf profile with a standard achiral GC column. Low amounts of 7S and 7R-cis-cis nepetalactones (4a and 4a’, respectively) were also observed in *N. sibirica* leaves. This result is surprising given that only 7S-specific ISY have been identified in *N. sibirica*, and throughout the Nepeta genus. We also tested all *N. sibirica* NEPS for ISY activity, but observed no activity for this reaction. Assays of 8-oxoneral 6 (which would result from neryl-xyrophosphate intermediate instead of geranyl xyrophosphate) with NEPS or NSP58 was also unreactive toward 7R products (Supplementary Fig. 11). It is likely that the *N. sibirica* 7R-ISY is not a homolog of ISY or NEPS and therefore identification of this enzyme will require a more intensive investigation. In summary, the high expression and in vitro activity of NSNEPS1A strongly suggests that this enzyme is a bona fide trans-trans cyclase and oxidase that is responsible for the synthesis of 7R-trans-trans nepetalactone 4d in planta.

Combinatorial biosynthesis of nepetalactol and nepetalactone isomers

Enzymes that can catalyze formation of the four 7S stereoisomers were now in hand: (i) 7S-cis-trans (7S,4aS,7aR) with NSNEPS2 (lactone 4a) and NSNEPS2-Y167F (lactol 3a); (ii) 7S-cis-cis (7S,4aR,7aS) 4b with NmNEPS3-Q206V or NcNEPS3A; (iii) 7S-trans-trans (7S,4aS,7aS) 4c with NmNEPS1-155SVTA; (iv) 7S-trans-cis (7S,4aR,7aR) 4d with NsMLPL2 (albeit as a minor component). Furthermore we showed that we can access 7R-trans-trans (7R,4aS,7aS) nepetalactone 4d' using 7R-ISY and NSNEPS1A/B. We set out to determine whether we could generate the other 7R stereoisomers using combinations of NEPS and 7R-ISY to gain access to diverse iridoid stereoisomers.

Previous work had established that 7R-ISY produces 7R-cis-trans (7R,4aR,7aS) nepetalactone 3a' through spontaneous cyclization, though numerous iridodial byproducts are observed. NSNEPS2, 7S-cis-trans nepetalactone 4a synthesis when assayed with 7S-ISY, yielded 7R-cis-trans nepetalactone 4a' when assayed with 7R-ISY. Correspondingly, NSNEPS2-Y167F yielded 7R-cis-trans nepetalactone 3a' (Fig. 6A, B). Assays of the remaining NEPS with 7R-ISY revealed that: NSNEPS4produced trace levels of 4a' while no activity was observed for NSNEPS4A. NSNEPSL had 4a' activity though low levels of 7R-trans-trans nepetalactone 4d' were also observed (Supplementary Fig. 12). NEPS3-Q206V produced 7S-cis-cis (7S,4aR,7aS) nepetalactone 4b (Fig. 2B), and as a result we hypothesized that this enzyme could also generate the enantiomer 7R-cis-cis nepetalactone 4b'. However, NmNEPS3-Q206V did not produce 4b' as expected but 3a' (Fig. 6B). Analogously, we attempted to use NmNEPS3-Q155SVTA (Fig. 3B), which produced 7S-trans-cis 4c, but instead we observed 7R-trans-trans (7R,4aS,7aS) 4d' nepetalactone when assayed with 7R-ISY (Fig. 6D). Moreover, the fraction of 7R-trans-trans nepetalactone 4d' relative to 7R-cis-trans nepetalactone 4a' when NmNEPS1-155SVTA is employed appears to be larger than NSNEPS1A or NSNEPS1B. Screening of 7R ISY with all individual NEPS and MLP enzymes failed to produce an enzyme system that yielded 7R-cis-cis 4b' or 7R-trans-cis 4c'. Finally, we explored the possibility of using the more pro-nicous NSNEPS1B in combination with *N. sibirica* MLP enzymes. Although NSNEPS1B does not produce 7R-cis-cis nepetalactone when combined with NSMLPL2, 7R-cis-cis 4b' appears as one of the secondary products (Fig. 6C). Importantly, although enzymic yields of 4b' could be further optimized, 4b' can be readily obtained from enzymatically derived 7R-trans-trans nepetalactone 4d' via chemical epimerization (Fig. 5C). Therefore, 7 out of the 8 possible nepetalactone stereoisomers have been accessed in vitro using a combination of native and mutant NEPS/MLPLs along with the two stereo-selective ISY enzymes. In summary, the active site of NSNEPS2 can accommodate formation of both enantiomers, 7R-cis-trans 4a and 7S-trans-trans 4a' which have two unique 4a-7a bridge configurations (4aR,7aS vs. 4aS,7aR, respectively). Similarly, NSNEPS1B can also accommodate formation of both enantiomers, 7R-trans-trans 4d and 7R-trans-trans 4d' through formation of 4d' is clearly favored. In contrast, NmNEPS3-Q206V and NmNEPS3-155SVTA appear to be highly specific for the bridgehead carbon stereochemistry (4aR,7aS and 4aS,7aS, respectively) regardless of the C7 methyl group configuration (Fig. 6E).

Finally, we surveyed additional *N. mussinii*, *N. cataria* and *H. officinalis* selected NEPS and MLP enzymes. Surprisingly, *H. officinalis*, a mint family plant that does not accumulate nepetalactones, possesses NEPS-like (NEPSL) enzymes that are able to oxidize both 7S and 7 R cis-trans nepetalactol (3a and 3a', respectively) to form the corresponding nepetalactones (Supplementary Fig. 13). Moreover, NEPSL enzymes of *N. mussinii*, *N. cataria* and *N. mussinii* NEPS5 enzymes, which do not have cyclization activity for 7S substrates do catalyze cyclization of 7R-trans-trans nepetalactol 3d' and its subsequent oxidation to nepetalactone 4d'. Taken together, these results suggest that nepetalactol oxidation activity is more widespread in the mint family than previously thought and that it is perhaps the context of the presence or absence of an ISY-like enzymatic activity that controls whether the plant accumulates nepetalactones, or whether it is of 7S or 7R configuration.

The iridoid scaffold has potential applications as an insect repellent but it is difficult to chemically synthesize in a stereo-selective manner. Here we report the discovery of NEPS enzymes, both naturally occurring and engineered, capable of stereo-divergent enzymatic synthesis of 7 iridoid scaffold stereoisomers. Gene discovery efforts in a phylogenetically distant *Nepeta* species led to the discovery of a NEPS with previously uncharacterized cyclase activity. Structural characterization and mutagenesis demonstrated that cyclization and oxidation activity could be decoupled in the NEPS, resulting in the production of products that were not observed in the natural plant extracts.
in more versatile biocatalysts. Moreover, structural elements that controlled the stereo-selectivity of the cyclization were identified in the NEPS proteins, providing mechanistic insight into this unusual class of cyclases. The discovery of a group of naturally occurring orthologues with varying product selectivity served as a starting point for modulating oxidative and cyclization selectivity. The impact of the engineered mutations was often dependent on the specific protein. While this precluded formulating a generalizable set of design rules for engineering these enzymes, we nevertheless could make some hypotheses regarding the mechanistic basis of how these mutations modulated cyclization and oxidation activities. Finally, we could productively combine cyclases to lead to a non-naturally occurring stereoisomer.

Plants have evolved an expansive chemical repertoire comprising \(10^5\) to \(10^6\) known compounds\(^1\). Our ability to understand and harness this chemical diversity has wide ranging applications in agricultural pest management, crop domestication, food industry, medicines, and specialty chemical production. Thus, identification and engineering of the biosynthesis genes that are responsible for the synthesis of these molecules presents opportunities for...
production of high value molecules using metabolic engineering and synthetic biology approaches. To date, 7S-cis-trans nepetalactol 3a, which is also formed spontaneously in the absence of any cyclase, albeit in greatly reduced yield, has been the main focus of microbial engineering efforts. The enzymes reported in this study can be utilized in these metabolic engineering systems to expand the iridoid-derived natural product space and potentially generate natural product derivatives.

Methods

Chemicals

8-oxogeranial (CAS# 80054-40-6) & 7S-cis-trans nepetalactol (CAS# 109215-55-6) from Toronto Research Chemicals. 7S-cis-trans and 7S-trans-cis nepetalactone were isolated from Nepeta fassinii "Six Hills Giant" (Burncoose Nurseries, Gwennap, UK) and from catnip oil, respectively using flash silica gel chromatography. 7S-cis-cis nepetalactone was isolated from Nepeta mussinii "Snowflake"
Cloning
Primers and synthetic genes were purchased from Integrated DNA Technologies (IDT). To clone NEPS, MLPL and ISY genes from *N. sibirica* cDNA library, primers (see Supplementary Table 2) contained overhangs with homology to the pOPINF vector (Addgene plasmid #26042). Putative genes from *N. sibirica* as well as all NEPS variants generated, were cloned into the pOPINF vector using the commercial kit Infusion HD (Clontech). Extracted RNA (see transcriptome section) was DNase treated with TURBO DNA-free kit (Thermo Fischer) and used to build cDNA library using SuperScript IV VILO (Invitrogen) master mix. Mutations to NEPS genes were done with primers containing the sequence as published data43. DNA fragments with overhangs with homology to the sequences of the Lamiaceae ISY orthogroup from previously published data. DNA fragments with overhangs with homology to pQLinkH were ordered from Thermofisher GeneArt Strings service, and cloned into the pQLinkH vector (Addgene plasmid # 13667) using the commercial kit Infusion HD (Clontech).

Enzyme expression and purification
Unless otherwise noted, genes cloned into pOPINF vectors (NEPS, MLPL, ISY) and pQLinkH (LaSYY) vectors were transformed into *E. coli* strain for protein expression and carbencillin (50 µg/mL) was used for selection. Single colonies grown in LB agar media were picked and grown overnight in 10 mL LB media at 37 °C. Next day, 1 mL of the overnight cultures were used to inoculate 100 mL of 2xYT media. The fresh cultures were grown at 37 °C until OD600 of 0.6-0.8 followed by induction with 250 µM IPTG and incubation at 18 °C for 16-18 h. For purification, the cells were harvested by centrifugation (10 min at 3200 g), resuspended in 10 mL of binding buffer (50 mM tris-HCl pH 8, 50 mM glycine, 500 mM Sodium Chloride, 250 mM imidazole, 5% v/v glycerol, pH 8) containing in addition 0.2 mg/mL Lysozyme and EDTA free protease inhibitor cocktail (Roche cOmplete). After incubation on ice for 30 min, the cells were lysed by sonication using a Sonics Vibra Cell at 40% amplitude, 3 s ON, 2 s OFF, 4 min total. Next, the lysate was centrifuged at 35,000 g for 15 min and the cleared lysate was applied to Akta pure system for purification. The purification was done in three steps. First the lysate as applied to a 5 mL HiTrap column (GE) to bind NEPS and eluted with a buffer containing 500 mM imidazole. Next the eluted protein was applied to a HiLoad 16/60 Superdex 200 pg (GE) with 20 mM HEPES pH 7.5, 150 mM NaCl as running buffer for size exclusion purification. Finally, the protein was concentrated and buffer exchanged into Low salt buffer (20 mM Tris-HCl pH 7.5, 200 mM NaCl) for ion exchange purification. The protein was applied to a Mono Q 5/50 GL (GE) column and eluted with High salt buffer (20 mM Tris-HCl pH 7.5, 1000 mM NaCl) using a linear gradient. The protein was concentrated and buffer exchanged back into Low salt buffer using Amicon 10 kDa MWCO centrifuge filters and the final concentration of protein was 9.6 mg/mL.

Enzyme assays and extraction
End-point enzymatic assays where 8-oxogeranil is combined with ISY and NEPS/MLP/L were done under the following conditions unless otherwise noted; 50 mM HEPES pH 8, 0.5 mM 8-oxogeranil, 1 mM NADPH, 5 mM NAD+, 0.5 µM for CrI SY or 2 µM for LaSYY, and 10 µg of NEPS or MLPL in 100 µL total volume. The assays were set up in 1.5 mL Eppendorf tubes and incubated for 3 h at 30 °C at 700 rpm. It is important to ensure that NEPS protein stocks do not warm to room temperature unless NAD+ is present, otherwise, protein precipitation can occur. After reaction, 10 µL of 1 mM Camphor in acetonitrile was added to the reaction products as internal standard and the reaction was then extracted with 100 µL of ethyl acetate. The mixture was vortexed for 2 min and centrifuged at 15,000 g for 5 min before collecting the ethyl acetate layer for GC-MS analysis.

GC-MS analysis
For analysis of nepetalactones on achiral column, samples were injected into a Thermo Trace 1310 GC-MS with an ISQ LT mass selective detector and a CTC Analytics PAL GC-xt autosampler. The samples (1 µL) were injected at inlet temperature of 230 °C in 1:10 ratio split mode into a Zebron ZBS-HT-INFERNO (ID = 0.25 mm; L = 30 m; film thickness = 0.1 µm). The mobile phase was helium at constant flow of 1.1 mL/min. The oven temperature ramp was as follows: starting at 80 °C, hold 5 min, ramp at 2.5 °C/min to 110 °C, ramp at 50 °C/min to 280 °C, hold 4 min. The transfer line to MS was kept at 280 °C and the MS source was at 250 °C. After a solvent delay of 4 min, mass range of 50-330 amu was collected at a fragmentation energy of 70 eV.

For analysis of nepetalactones on chiral column, samples were injected into an Agilent 8890 GC system with an Agilent 5977B mass selective detector and an Agilent 7693A autosampler. The samples (1 µL) were injected at inlet temperature of 220 °C in L10 ratio split
mode into a Supelco Beta DEX 225 column (ID = 0.25 mm; L = 30 m; film thickness = 0.25 µm). The mobile phase was helium at constant flow of 11 mL/min. The oven temperature ramp was as follows: starting at 80 °C, hold 3 min, ramp at 10 °C/min to 120 °C, hold 45 min, ramp at 10 °C/min to 200 °C, hold 2 min. The transfer line to MS was kept at 220 °C and the MS source was at 230 °C. After a solvent delay of 10 min, mass range of 50–350 amu was collected at a fragmentation energy of 70 eV.

Phylogenetic tree and sequence alignments
Protein sequence alignment was constructed in Geneious Prime (2019.1.1) software using the ClustalW (v2.1) (Cost matrix = BLOSUM, gap open cost = 10, gap extend cost = 0.1) option. Phylogenetic tree was inferred with the Maximum Likelihood method and Poisson correction model tree was constructed in MEGAX 10.2.2 software using the Bootstrap test with 1000 replications, and a discrete Gamma distribution (5 categories).

Circular dichroism
For circular dichroism measurements, nepetalactone isomer samples were diluted in methanol at 0.1 mg/mL. Samples at room temperature (20 °C) were placed in a cuvette with cell length of 1 mm. Spectra were measured on a Jasco J-810 spectropolarimeter from 400–190 nm in 1 nm steps at a scanning speed of 100 nm/min.

NsNEPS2 crystallization
NsNEPS2 was purified as described above. NsNEPS2 crystals were formed using the sitting drop method. The precipitant solution was 0.2 M Sodium acetate with 11.5% w/v PEG 4000. Droplet conditions were: 0.6 µL total volume, 200 µL protein + 400 µL precipitant, 10x molar excess NAD+ in the protein stock. Prior to harvesting, crystals were soaked for 5 min in the precipitant solution containing 1.66 mM NAD+, 6.66 mM cis-trans-nepetalactone and 25% ethylene glycol as cryo-protectant. Data collected contained NAD+ bound but not nepetalactone.

NsNEPS2 crystals data collection and structure determination
X-ray data sets were recorded on the I03A (PX II) beamline at the Paul Scherrer Institute (Villigen, Switzerland) at wavelength of 1.0 Å using a Dectris Eiger 16 M detector with the crystals maintained at 100 K by a cryocooler. Diffraction data were integrated using XDS (BUILT = 202022020) and scaled and merged using AIMLESS (0.7.7); data collection statistics are summarized in Supplementary Table 1. Initially the NsNEPS2 data set was automatically processed at the beamline to 1.85 Å resolution and a structure solution was automatically obtained by molecular replacement using pdb 6F9Q as template. The map was of sufficient quality to enable 90% of the residues expected for a NsNEPS2 homodimer to be automatically fitted using Phenix autobuild (1.19.2). The model was finalized by manual rebuilding in COOT (0.9.6) and refined using in Phenix refine (1.19.2) and is deposited as PDB entry 7QUJ.

Protein crystal modeling and molecular docking
Homology models were built using the Swiss-Model server. When models were based on NmNEPS3 crystal structure, PDB 6F9Q was used as template. When models were based on NsNEPS2 (now PDB 7QUJ), the crystal structure map was introduced manually. Protein crystal modeling and molecular docking of nepetalactones into NEPS was performed using AutoDock Vina (1.1.2). Ligand 3D structures were generated in MarvinSketch (19.19.0) and AutoDockTools (1.5.6) was used to prepare the ligands and receptor molecules for docking.

N. sibirica genes transcriptome assembly and gene expression abundance estimation
Plants were grown in a greenhouse exposed to the temperature and humidity of the environment in Norwick, UK. Plants were cut back and young leaf tissue collected after 2 weeks. RNA was isolated using a CTAB method prior to DNase treatment using the Turbo DNA-Free Kit (ThermoFisher Scientific, Waltham, MA, Cat #AM1907), and checked for quality with a NanoDrop and a TapeStation (Agilent, Santa Clara, CA). RNA was then used to construct an Illumina TruSeq Stranded mRNA library (Illumina, San Diego, CA) and sequenced to 150 nt on an Illumina HiSeq4000 (Illumina, San Diego, CA) in paired end mode at Michigan State University’s Research Technology Support Facility. Reads were trimmed for adapter sequences and low-quality bases using Trimomatic (v0.32) and cleaned reads were normalized and used to generate strand-specific transcriptome assemblies using Trinity (v20140717) with default parameters, of which, only transcripts >200 bp were retained for subsequent analyses. The longest isoform for each transcript was deemed the representative transcript. A nucleotide BLAST database was generated based on the N. sibirica transcriptome and genes encoding for putative NEPS, ISY and MLPL were identified using TBLASTN with N. cataria and N. mussinii enzymes as the query. For the 7Rs ISY search, L. album ISY was also used as a query.

To generate expression abundances, plants were grown under greenhouse conditions (day temp: 22 ± 1 °C, night temp: 20 ± 1 °C, light cycle of 14 h, RT of 50% ± 5%) and replicated tissue was collected from immature leaves, mature leaves, green stems, underground stems, and flowers. RNA was isolated and prepared as above prior to sequencing on an Illumina HiSeq4000 (Illumina, San Diego, CA) to 50 nt in single end mode at Michigan State University’s Research Technology Support Facility. Replicated RNAseq reads were trimmed using Cutadapt (v3.4) removing bases from each end with quality scores <30 and employing a minimum read length of 30nt. Trimmed reads were then pseudo-aligned to the generated representative transcriptome and sequences of interest to generate Transcripts per Million (TPMs) using Kallisto (v0.46.2) in stranded mode employing the following two parameters: -l 200 -s 20. Reads are available in the National Center for Biotechnology Information under BioProject ID PRJNA762925.

Additional software
Additional software employed in this study are as follows. Assay graphs were made in GraphPad Prism (8.2.0). Adobe Illustrator (15.0.0) was used to generate final figures. Nucleic acid and protein sequences were analyzed/designed in Geneious Prime (2019.1.1). Chemical structures were generated in ChemDraw Professional (17.1), unless otherwise mentioned. ESPript3.0 online tool was used to generate the alignment figure. Protparam (ExPasy) online tool was used to calculate MW and extinction coefficients for determining protein concentrations.

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The RNA-seq data for N. sibirica generated in this study have been deposited in the National Center for Biotechnology Information under BioProject ID PRJNA762925. The NsNEPS2 crystal structure data generated in this study have been deposited in the Protein Data Bank (PDB) under accession code 7QUJ.

The characterized genes have been deposited in GenBank under accession codes: NsNSPSL, ON010447; NsNSPSIA, ON010448; NsNSPSIB, ON010449; NsNSPS2, ON010450; NsNSPS4A, ON010451; NsNSPS4B, ON010452; NsMLPL1, ON010453; NsMLPL2, ON010454; NsMLPL3, ON010455; NsISY, ON010456; NsPSBR, ON010457; LaISY, ON010458. Additionally, data is available from the corresponding authors upon request.
Mohamed Omar Kamileen for isolating trans-trans nepeta lactone from N. sibirica and helpful discussions on in-vitro assays. RNA extraction and general procedures. We thank Delia Serna for her assistance with GC-MS equipment and Dr. Stefan Bartram for his assistance with the CD-spectrometer equipment. This work was funded by a grant from the National Science Foundation Plant Genome Research Program (IOS-144449), the European Research Commission (788301) and the Max Planck Society. B.H. is supported by a postdoctoral fellowship from the Alexander von Humboldt Foundation.

Author contributions
Author contributions N.J.H.L. and S.E.O’C. designed the study and wrote the manuscript; N.J.H.L. performed experiments. B.H. performed chemical synthesis and NMR. J.C.W. and D.Z. assembled the transcriptome; J.B., L.C., and C.L. solved the crystal structure of NsNEPS2; L.Ch. assisted with cloning and assays experiments; M.K. assisted with GCMS; C.E.R.L. discovered LaISY; B.R.L. assisted with design of the study; C.R.B. provided bioinformatic and sequencing support.

Funding
Open Access funding enabled and organized by Projekt DEAL.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-32414-w.

Correspondence and requests for materials should be addressed to Sarah E. O’Connor.

Peer review information Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022

Acknowledgements
We would like to thank Eva Rothe for her assistance in the growth and maintenance of the Nepeta plants. We would also like to thank...