

Quality-Diversity Meta-Evolution: customising behaviour spaces to a meta-objective

David M. Bossens and Danesh Tarapore

Abstract—Quality-Diversity (QD) algorithms evolve behaviourally diverse and high-performing solutions. To illuminate the elite solutions for a space of behaviours, QD algorithms require the definition of a suitable behaviour space. If the behaviour space is high-dimensional, a suitable dimensionality reduction technique is required to maintain a limited number of behavioural niches. While current methodologies for automated behaviour spaces focus on changing the geometry of the behaviour space or on unsupervised learning of its key features, there remains a need for customising behavioural diversity to a particular meta-objective specified by the end-user. In the newly emerging framework of QD Meta-Evolution, or QD-Meta for short, one evolves a population of QD algorithms, each with different algorithmic and representational characteristics, to optimise the algorithms and their resulting archives to a user-defined meta-objective. Despite promising results compared to traditional QD algorithms, QD-Meta has yet to be compared to state-of-the-art behaviour space automation methods such as Centroidal Voronoi Tessellations Multi-dimensional Archive of Phenotypic Elites (CVT-MAP-Elites) and Autonomous Robots Realising their Abilities (AURORA). This paper performs an empirical study of QD-Meta on function optimisation and multi-legged robot locomotion benchmarks. Results demonstrate that QD-Meta archives provide improved average performance and faster adaptation to a priori unknown changes to the environment when compared to CVT-MAP-Elites and AURORA. A qualitative analysis shows how the resulting archives are tailored to the meta-objectives provided by the end-user.

Index Terms—quality-diversity algorithms, meta-evolution, evolutionary robotics

I. INTRODUCTION

Historically, most evolutionary algorithms (EAs) optimised a fitness function without considerations for generalisation to unseen problems or robustness to perturbations to the evaluation environment. However, it was widely known that successfully converging to the maximum of that fitness function requires maintaining genetic diversity in the population of solutions (e.g., [1]–[4]). Moreover, the use of niching demonstrated how maintaining subpopulations could help find multiple solutions to a single problem [5]. Some studies included genetic diversity as one of the objectives of the EA [6]. Approaches in evolutionary robotics, artificial life, and neuro-evolution realised that genetic diversity does not necessarily imply a diversity of solutions, since (i) different genotypes may encode the same behaviour and vice versa; and (ii) many genotypes may encode unsafe or undesirable solutions that should be discarded during evolution (e.g., when a robot crashes into an obstacle). Such approaches

began to emphasise *behavioural diversity* [7]–[10], not only as a driver for objective-based evolution but also as the enabler for diversity- or novelty-driven evolution [11].

In *quality-diversity* (QD) algorithms such as MAP-Elites [12] and Novelty Search with Local Competition [13], the behavioural diversity approach is combined with local competition such that the best solution for each local region in the behaviour space is stored, forming a large archive of high-quality solutions. The development of quality-diversity algorithms has allowed a plethora of applications. In robotics, this includes the design of robot morphologies and controllers [12], [14], as well as providing the behaviour space for behaviour adaptation algorithms [15], [16], which search for high-performing behaviours in the evolved archive to help robots recover rapidly from environmental changes or damages to their sensory-motor systems.

A growing number of approaches has started to explore automatically defined behaviour spaces for QD algorithms. For Novelty Search, Meyerson et al. (2016) [17] proposed to learn a behavioural distance function to suit a particular domain of problems by increasing weights of those behavioural features that are critical to success on the target domain. For MAP-Elites, low-dimensional behaviour spaces are particularly desirable since the number of niches grows exponentially with the number of dimensions in the behaviour space. In CVT-MAP-Elites [18], Centroidal Voronoi Tessellations provide a different geometry to MAP-Elites such that each niche is defined by the regions around a centroid in the behaviour space. Since the user can pre-specify the number of centroids, this avoids the exponential increase in the number of niches. In AURORA [19], standard dimensionality reduction algorithms such as Principal Component Analysis (PCA) and auto-encoders are applied to automating the behaviour space in a distance-based archive as in Novelty Search with Local Competition. In a first study, AURORA outperformed CVT-MAP-Elites and also demonstrated the benefit of auto-encoders over PCA [19]. A follow-up study [20] further showed the fitness advantage over TAXONS [21], a comparable behaviour space automation method which also uses auto-encoders but which emphasises diversity but not quality.

In a variety of applications, end-users may be particularly interested in behaviour spaces that are custom-made to a particular *meta-objective*, which expresses desirable properties for the final archive, such as a high number of solutions, generalisation towards particular problems, ability to form meaningful behavioural sequences, etc. In this context, we propose the newly emerging framework of *quality-diversity meta-evolution* [22], [23], or QD-Meta for short, to evolve a

population of QD algorithms, each with their own behaviour space and optionally some other representational or algorithmic properties. The framework learns a feature-map to define the behaviour space and through the use of a large database of promising solutions generated so far, it allows to rapidly populate new archives based on the adapted behaviour space. The proposed approach also controls dynamically, based on the meta-population statistics, how many generations the QD archives will evolve before they are evaluated on the meta-objective. Both aspects have been studied in prior work, with the evolution of feature-maps yielding large generalisation improvements compared to MAP-Elites variants [22], [23] and with parameter control of the generations yielding small but consistent benefits in QD metrics and in the meta-objective [23]. This paper aims to further demonstrate QD-Meta as a unique framework for QD optimisation by investigating the following research questions:

- How does QD-Meta compare to state-of-the-art methods AURORA and CVT-MAP-Elites on standard QD metrics as well as on behaviour adaptation to dynamic changes to the fitness landscape, for instance, due to changes in the environment?
- What are the unique properties of the archives, in terms of quantitative metrics but also qualitative behaviour characterisations, as they are tailored to different meta-objectives?
- What is the benefit of the population-based methodology of QD-Meta, in which multiple feature-maps are developed, and how does the meta-population size affect QD and meta-fitness measures of QD-Meta?
- Previous QD-Meta work has used feature-sets, which are sets of behavioural descriptors selected by the user. Are such hand-crafted feature-sets a requirement for QD-Meta's performance?

These questions are investigated on two distinct benchmarking domains of high importance for the evolutionary computation community, namely Rastrigin function optimisation and multi-legged robot locomotion.

II. QUALITY-DIVERSITY META-EVOLUTION

This section details our quality-diversity meta-evolution algorithm. In this study it is implemented using Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) [24], [25] to evolve a population of MAP-Elites [12] algorithms (see Fig. 1 for an illustration).

A. MAP-Elites algorithm

The MAP-Elites (ME) algorithm [12] discretises the behaviour space into behavioural bins, which are equally-sized hypercubes, and then maintains for each behavioural bin the elite solution (i.e., the solution with the highest fitness), leading to quality-diversity.

ME first randomly generates an initial population of genotypes. Then, each genotype in the initial population is evaluated, resulting in a fitness score f and a behavioural descriptor β . Each genotype is then added to the behaviour-performance map \mathcal{M} based on the following elitist replacement rule: if

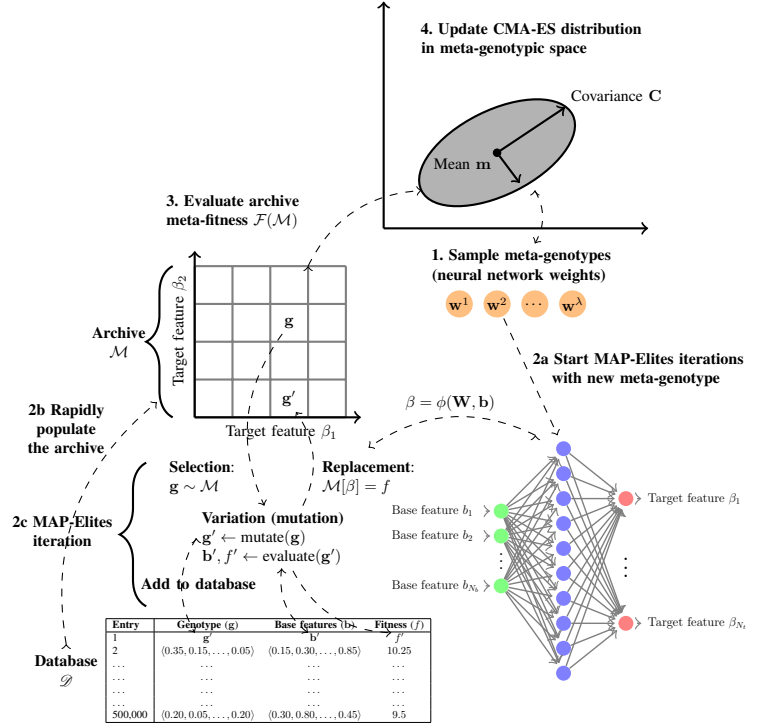


Fig. 1. Flow diagram of QD-Meta, which repeats a four-step cycle: **1.** CMA-ES samples the meta-population from its distribution in the meta-genotypic space. **2.** Each of the meta-genotypes w^i for $i \in \{1, \dots, \lambda\}$ then independently applies MAP-Elites with its own feature-map: **2a** construct the feature-map $\phi(w^i, \cdot)$; **2b** rapidly populate the archive \mathcal{M}^i with entries from the database \mathcal{D} . **2c** perform repeated MAP-Elites iterations, selecting a genotype from the archive, mutating it, evaluating its fitness f' and its base-features b' from the observed behaviour, computing the target-features $\beta = \phi(w^i, b')$, adding the solution to the database, and adding it in the archive as $\mathcal{M}[\beta]$ – if it is an elite for the region around β . **3.** The archives evolved by the meta-genotypes are evaluated on their meta-fitness. **4.** Using the meta-genotypes and their meta-fitness scores, CMA-ES updates its distribution in the meta-genotypic space to find archives with the highest meta-fitness.

the genotype is the best solution so far for that behavioural bin – either because β is empty ($\mathcal{M}[\beta] = \emptyset$) or because the fitness is higher than the current genotype in that bin ($f > f(\mathcal{M}[\beta])$) – then store that genotype g in that bin of the behaviour-performance map (i.e., $\mathcal{M}[\beta] \leftarrow g$).

After initialisation, the algorithm applies repeated cycles of random selection, genetic variation, evaluation, and replacement. Random selection is implemented by randomly selecting genotypes from non-empty behavioural bins in the behaviour-performance maps. Genetic variation is based on mutations to the genotypes. Evaluation of genotypes is based on a user-defined fitness function $f(\cdot)$. Replacement is based on the above-mentioned elitist replacement rule. After many repetitions of this cycle, applying mutations on genotypes selected randomly from a growing set of filled behavioural bins ultimately leads to covering most behavioural bins of the behaviour-performance map. Due to repeatedly replacing existing genotypes with higher-performing genotypes, the fitness potential of each covered behavioural bin will be illuminated.

B. QD-Meta algorithm

There are two key limitations to ME: 1) it is not suited to high-dimensional behaviour spaces; and 2) since the elites in the resulting behaviour-performance maps are not at any time evaluated in changed environments, they are not necessarily selected for their generalisation potential. Quality-diversity meta-evolution provides a promising perspective to overcome these issues by using feature-maps to formulate low-dimensional behaviour spaces and by evaluating its elites on a meta-objective that includes changed environments.

The QD-Meta algorithm implementation uses Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) [24], [25] to evolve a population of MEs – the meta-population – and thereby optimise behaviour-performance maps with generalisation as a meta-objective. The algorithm, described in detail in Algorithm 1, first applies an initialisation phase to populate the behaviour-performance maps in which a large number of random genotypes are sampled, evaluated, and then added to the database \mathcal{D} (see l. 1–6, and Section II-C for further details on the database). After initialisation, the meta-evolution algorithm performs a large number of *meta-generations*, consisting of the following steps:

- 1) Sample new meta-genotypes $\mathbf{W}^1, \dots, \mathbf{W}^\lambda$ from the multivariate normal distribution defined by CMA-ES (see l. 10).
- 2) For $i \in \{1, \dots, \lambda\}$, use meta-genotype \mathbf{W}^i to construct a new map \mathcal{M}^i based on existing solutions in the database (l. 11–12).
- 3) Get the meta-population state s' and select the number of ME iterations a' based on the Q -table defined by SARSA (l. 15–16).
- 4) For $i \in \{1, \dots, \lambda\}$, ME($a', \mathcal{M}^i, \mathbf{W}^i$) further evolves \mathcal{M}^i (l. 18 and l. 36–42). Newly generated solutions, including their genotype, base-features, and fitness, are stored in the database to be used in step 2) of the next meta-generation.
- 5) Evaluate each meta-genotype $i \in \{1, \dots, \lambda\}$ on the meta-fitness $\mathcal{F}(\mathbf{W}^i)$ (l. 19; see Section III for its definition).
- 6) SARSA(λ) uses the new experience $\langle s, a, r, s', a' \rangle$ to update Q (see Section II-E).
- 7) CMA-ES updates the mean, covariance, and step size, applying the $(\mu/\mu_W, \lambda)$ -CMA Evolution Strategy [25] (l. 23–25).

C. Database

The database \mathcal{D} stores a large number of previously found solutions to enable rapidly generating new behaviour-performance maps. Each such solution is a tuple $\langle \mathbf{g}, \mathbf{b}, f \rangle$, where \mathbf{g} is the genotype (e.g., the parameters of a controller), \mathbf{b} is the base-behavioural description of the solution according to a large number of N_b user-defined features, and f is the fitness (e.g., the performance of a controller).

Two database types have been explored in prior work, a circular buffer [22] and the k -best database designed to preserve high-performing solutions with high behavioural diversity [23]. We opt for the former, as the latter is not scalable to high-dimensional base-behavioural spaces.

Algorithm 1 QD-Meta

```

1:  $\mathcal{D} \leftarrow \emptyset$ . # Create empty database.
2: for  $i = 1$  to  $p$  do # Create initial database.
3:    $\mathbf{g} \leftarrow \text{random-genotype}()$ .
4:    $\mathbf{b}, f \leftarrow \text{eval}(\mathbf{g})$ . # Base-features and fitness.
5:   Insert  $\langle \mathbf{g}, \mathbf{b}, f \rangle$  into  $\mathcal{D}$ . # Fill the database (see Section II-C).
6: end for
7: for  $j = 1$  to  $G$  do # Loop over meta-generations.
8:   for  $i = 1$  to  $\lambda$  do
9:     Set  $\mathcal{M}^i \leftarrow \emptyset$ . # Empty the map.
10:     $\mathbf{w} \sim \mathcal{N}(\mathbf{m}, \sigma \mathbf{C})$ . # Sample meta-genotype.
11:    for  $\langle \mathbf{g}, \mathbf{b}, f \rangle \in \mathcal{D}$  do # Construct map from database.
12:      add-to-map( $\mathcal{M}^i, \mathbf{w}, \mathbf{g}, \mathbf{b}, f$ ).
13:    end for
14:  end for
15:  Get meta-population state  $s'$ . # see Section II-E
16:   $a' \leftarrow \text{epsilon-greedy}(s')$ . # Control ME iterations.
17:  for  $i = 1$  to  $\lambda$  do
18:    Perform ME-iterations( $a', \mathcal{M}^i, \mathbf{W}^i$ ).
19:     $\mathcal{F}_i \leftarrow \text{Meta-fitness}(\mathcal{M}^i)$ .
20:  end for
21:  SARSA( $\lambda$ ) update on  $\langle s, a, r, s', a' \rangle$ . # see Eq. 2
22:   $s \leftarrow s'; a \leftarrow a'$ .
23:   $\mathbf{m} \leftarrow \text{Update-mean}()$ .
24:   $\mathbf{C} \leftarrow \text{Update-covariance}()$ .
25:   $\sigma \leftarrow \text{Update-step}()$ .
26: end for
27: procedure ADD-TO-MAP( $\mathcal{M}, \mathbf{w}, \mathbf{g}, \mathbf{b}, f$ )
28:    $\mathbf{W} \leftarrow \text{transform}(\mathbf{w})$ . # Transform meta-genotype
29:   # (see Section II-D).
30:    $\beta \leftarrow \phi(\mathbf{W}, \mathbf{b})$ . # Apply feature-map to get target features
31:   # (see Eq. 1).
32:   if  $\mathcal{M}[\beta] = \emptyset$  or  $f > f(\mathcal{M}[\beta])$  then
33:      $\mathcal{M}[\beta] \leftarrow \mathbf{g}$ . # Add genotype  $\mathbf{g}$  to the map  $\mathcal{M}$ .
34:   end if
35: end procedure
36: procedure ME-ITERATIONS( $I, \mathcal{M}, \mathbf{w}$ )
37:   for  $i = 1$  to  $I$  do #  $I$  is the number of iterations.
38:      $\mathbf{g} \sim \mathcal{M}$ . # Sample genotype randomly from map.
39:      $\mathbf{g}' \leftarrow \text{mutate}(\mathbf{g})$ . # Mutation.
40:      $\mathbf{b}, f \leftarrow \text{eval}(\mathbf{g}')$ . # Base-features and fitness.
41:     add-to-map( $\mathcal{M}, \mathbf{w}, \mathbf{g}', \mathbf{b}, f$ ).
42:     Insert  $\langle \mathbf{g}', \mathbf{b}, f \rangle$  into  $\mathcal{D}$ . # Fill the database.
43:   end for
44: end procedure

```

D. Non-linear feature-maps

In QD-Meta, feature-maps are used to transform the base-behavioural features $\mathbf{b} \in [0, 1]^{N_b}$ to a low-dimensional behavioural descriptor $\beta \in [0, 1]^{N_t}$. We use non-linear feature-maps based on a feed-forward neural network with scaled sigmoid function,

$$\begin{aligned}
 n(\mathbf{W}, \mathbf{b}) &= \mathbf{W}^2 S_{N_b}(\mathbf{W}^1 \mathbf{b} + B^1) + B^2 \\
 \phi(\mathbf{W}, \mathbf{b}) &= S_{N_h}(n(\mathbf{W}, \mathbf{b})),
 \end{aligned} \tag{1}$$

where $S_N(\mathbf{x}) = 1 / (1 + \exp(-\alpha_s \mathbf{x} / (N + 1)))$ is an element-wise sigmoid function that scales the incoming activation based on the number of incoming units, N ; α_s is an empirically defined scaling factor; and the transformed meta-genotype, \mathbf{W} , is composed of a weight matrix from input to hidden layer, $\mathbf{W}^1 \in \mathbb{R}^{N_h \times N_b}$, a weight matrix from hidden layer to output, $\mathbf{W}^2 \in \mathbb{R}^{N_t \times N_h}$, and the corresponding bias units $B^1, B^2 \in \mathbb{R}$. For networks such as $\phi(\mathbf{W}, \mathbf{b})$, universal approximation theorems (e.g., [26], [27]) imply that in principle all multi-variate functions over closed and bounded intervals can be represented to arbitrary precision, assuming a sufficient number of neurons.

Such feature-maps strongly outperform other linear and feature-selector feature-maps in practice (see [23]). Beyond the ability to represent arbitrary input-output mappings, the output sigmoid activation in Eq. 1 also has a favourable statistical profile. It accounts for the high frequencies of near-zero values of $n(\mathbf{W}, \mathbf{b})$ as it increases steeply for values close to zero and slowly for extreme values. This ensures that for a sizeable proportion of mappings, each bin in $[0, 1]^{N_t}$ is frequently represented, leading to diversity and local competition in the target-feature space – the key requirements for quality-diversity.

E. Dynamic parameter control

An additional feature of the algorithm is dynamic parameter control to find high-performing dynamic schedules of parameters such as the number of generations per meta-generation and the mutation rate. Compared to other dynamic parameter control schedules, reinforcement learning (RL) using the SARSA(λ) implementation of Karafotias et al. (2014) [28] has been demonstrated as the method of choice in QD-Meta [23]. This paper will use the same RL technique for dynamic control of the number of generations per meta-generations to solve the trade-off between quality-diversity (base-generations) and the evaluation of the quality-diversity archive in terms of the meta-fitness (meta-generations). An advantage of RL over other adaptive schedules, such as annealing, endogenous control, multi-armed bandits, etc. (see [29] for annealing and endogenous control in QD and [30] for an overview), is that RL makes use of a rich observation history with various indicators of meta-evolutionary progress.

In the RL setup for QD-Meta parameter control, the algorithm uses intervals of the parameter setting as “actions” and then learns which actions result in meta-fitness improvements given observations on the meta-evolutionary progress. Observations are the maximum, mean, and standard-deviation of meta-fitness, the meta-genotypic diversity, the number of consequent meta-generations the maximal meta-fitness has not improved, and the reward. The reward is the ratio improvement in maximal meta-fitness divided by the function evaluations performed in the latest meta-generation. To ensure the state-space is not too large with such continuous observations, the state of the RL agent is formed using a tree-based discretisation [31], in which states represent different partitions with Q-values

indicated as significantly distinct according to a Kolmogorov-Smirnoff test.

The RL parameter control follows the algorithm by Karafotias et al. (2014) [28]. It uses the SARSA algorithm [32], [33], an on-policy version of Q-learning, with eligibility traces (see SARSA(λ) [33]). The algorithm performs ϵ -greedy action selection, selecting the best action according to $a^* = \arg \max_a Q(s, a)$ with probability $1 - \epsilon$ and selecting a random action with probability ϵ . After each experienced transition of state, action, reward, state, and action, $\langle s_t, a_t, r_t, s_{t+1}, a_{t+1} \rangle$, the Q-table is updated for all eligible state-action pairs as follows:

$$Q(s, a) \leftarrow \alpha e(s, a) (r_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s, a)) + (1 - \alpha) Q(s, a), \quad (2)$$

where γ is the discount factor (discounting the future rewards); α is the learning rate; and the eligibility $e(s, a)$ gives more weight to recently experienced state-action pairs, which addresses the temporal credit assignment problem in a way that works well for online learning in a non-episodic environments.² This is done by assigning $e(s_t, a_t) \leftarrow 1$ and $e(s, a) \leftarrow e(s, a) \lambda \gamma$ for all other state-action-pairs $(s, a) \in \mathcal{S} \times \mathcal{A}$.

III. EXPERIMENT SETUP

To illustrate the principles of QD meta-evolution on a variety of applications, experiments are conducted on two benchmarks, Rastrigin function optimisation and hexapod robot locomotion.³ In both benchmarks, there are two phases:

- 1) **Evolution phase:** the behavioural archives are evolved using QD algorithms; a summary of evolutionary parameters is given in Table I.
- 2) **Test phase:** the behavioural archives are assessed on fitness landscapes not experienced during evolution.

The key hypothesis is that if robustness to dynamic changes is incorporated to the meta-fitness, then archives evolved with QD meta-evolution will have improved performance on the test phase when compared to other QD algorithms.

A. Baseline algorithms

To evaluate QD meta-evolution compared to the state-of-the-art for low-dimensional behaviour spaces in QD, we compare to two baseline algorithms:

- **AURORA** [20]. To represent unsupervised automation of low-dimensional behaviour spaces, we include the popular AURORA algorithm mentioned in the introduction. Based on its superior overall performance over other AURORA variants, we choose AURORA-CSC-Uniform, which uses Container Size Control (CSC), a proportional control technique to match a desired pre-specified container size, and uniform selection over the archive. Out of the two neural network architectures for the feature-maps experimented with in the literature, Multi-Layer Perceptron (MLP) and Convolutional Neural Networks

²The dynamic parameter control problem is non-episodic since there is only a single lifetime with no terminal states

³Open-source code for replicating the experiments is available on the repository https://github.com/resilient-swarms/QDMeta_experiments.

¹Hidden units receive a weighted activation based on $N = N_b$ input units while output units receive a weighted activation based on $N = N_h$ hidden units

(CNN), we select the MLP as the base-features are not spatially correlated pixel data as is typical in CNN applications. The implementation is based on the source code provided in Grillotti & Cully (2021) [20] (see <https://github.com/adaptive-intelligent-robotics/AURORA>).

- **CVT-MAP-Elites** [18]. An alternative methodology for automated low-dimensional behaviour spaces, without use of feature-maps, is to formulate the problem in one dimension based on a few prototype behaviours. To represent this approach for dimensionality reduction in QD algorithms, we include Centroidal Voronoi Tessellations MAP-Elites, in which the niches are determined by a behaviour's distance to a limited number of pre-defined centroids spread evenly across the behaviour space. The implementation is based on the CVT-MAP-Elites module for the *sferes2* framework (see https://github.com/sferes2/cvt_map_elites), which was used in Vassiliades et al. (2021) [18].

In the experiments, the baseline algorithms and QD-Meta algorithms have the same base-behavioural features, denoted as $\mathbf{b} \in [0, 1]^{N_b}$. For AURORA and QD-Meta, base-features are projected onto a set of target-features, $\beta \in [0, 1]^{N_t}$ by means of their feature-maps. For CVT-MAP-Elites, the centroids are formulated within the same base-feature space.

TABLE I

PARAMETER SETTINGS FOR EVOLUTION. TOP HALF SHOWS SETTINGS COMMON TO ALL CONDITIONS WHILE BOTTOM HALF SHOWS SETTINGS FOR QD-META CONDITIONS.

Parameter	Setting for Rastrigin	Setting for RHex
Genotype (\mathbf{g})	$[0, 1]^{20}$	discretised in $[0, 1]^{24}$
Base-behaviour space	genotype	trajectory in $[0, 1]^{50}$ or feature-sets in $[0, 1]^{15}$
Mutation rate	0.10	0.125
Mutation type	Gaussian with SD $\sigma = 0.05$	random increment or decrement with step of 0.025
Maximal number of solutions	10,000 solutions	4,096 solutions
Function evaluations	100,000,000	12,000,000
Batch size per generation	400	400
Initial population (p)	2,000	2,000
Meta-population size (λ)	10	10
Number of target-features (N_t)	2	4
Number of hidden units (N_h)	10	10
Meta-genotype (\mathbf{w})	$[-1, 1]^{222}$	$[-1, 1]^{542}$
Sigmoid scaling factor (α_s)	30	30
Database capacity	500,000	500,000

B. Rastrigin function optimisation benchmark

In a first set of experiments, quality-diversity archives are evolved on a 20-dimensional Rastrigin function and then assessed on a set of test-perturbations, which represent a dynamic change of the fitness landscape. While standard genetic algorithms can solve the 20-D Rastrigin near-optimally, the archives evolved by QD algorithms have a unique benefit of covering various peaks of a fitness landscape; therefore, QD algorithms can potentially provide a near-optimal solution after a dynamic landscape change. The following two independent QD-Meta conditions are run, QD-Meta Dimension and QD-Meta Translation, corresponding to two different meta-objectives: 1) dimensionality decreases which remove one input-dimension from the Rastrigin computation; and 2) linear transformations

over the fitness landscape. The test phase runs two independent test scenarios, each of which are applied to all the algorithms. The Dimension test consists of injecting sinusoidal noise to selected pairs of dimensions while the Translation test consists of performing linear transformations. The resulting fitness landscapes of the tests are out-of-distribution even for the QD-Meta conditions.

a) *Evolution phase*: The 20-dimensional Rastrigin function is highly multi-modal and non-linear, posing a challenging optimisation problem. Formulated as a maximisation problem it takes the following form:

$$f(\mathbf{g}) = -10N_g \left(\sum_{i=1}^{N_g} g_i^2 - 10 \cos(2\pi g_i) \right), \quad (3)$$

where N_g is the number of genes in the genotype, $\mathbf{g} \in [-5.12, 5.12]^{N_g}$.

The following settings are common to all algorithms. The Rastrigin with its continuous search space is optimised by a Gaussian mutation operator with standard deviation of $\sigma = 0.05$ and a mutation rate of 0.10. As the Rastrigin experiments have low-cost function evaluations, a large budget of 100 million function evaluations and a large number of 10,000 solutions are allowed. The base-behavioural features of a solution are equal to its genotype.

Further algorithm-specific settings are as follows. For generating its centroids, CVT-MAP-Elites uses the k-means algorithm on randomly generated data points; here, we use 1 million data points of randomly generated points in $[0, 1]^{20}$. For QD-Meta, the meta-population size is increased to 10 compared to 5 in prior works [22], [23], mainly to discourage getting stuck in local optima of the meta-fitness. Two distinct QD-Meta conditions are included according to two different meta-objectives. In a first condition, called **QD-Meta Dimension**, the meta-fitness aims to provide behaviourally diverse and high-quality solutions despite changes to the dimensionality of the Rastrigin function. For QD-Meta Dimension, the meta-fitness of an archive \mathcal{M} combines the archive-summed fitness, averaged over different dimensionality decrements, and the summed pairwise distance in the genotypic space (here also the base-behavioural space):

$$\mathcal{F}_d(\mathcal{M}) = -\frac{1}{|\mathcal{J}|} \left(\sum_{j \in \mathcal{J}} \sum_{\mathbf{g} \in \hat{\mathcal{M}}} M + \bar{f}_{\{j\}}(\mathbf{g}) + \alpha \sum_{\mathbf{g}' \neq \mathbf{g}} \|\mathbf{g} - \mathbf{g}'\|_2 \right), \quad (4)$$

where $\mathcal{J} \subset \{1, \dots, 20\}$ is a set of 10 indices randomly selected before the start of evolution; $\hat{\mathcal{M}}$ is a random subset of 10% of the solutions of the full archive, selected anew for each meta-fitness evaluation; an upper bound M normalises $\bar{f}_j(\mathbf{g})$ to the range $[0, M]$ to obtain a positive score; $\bar{f}_{\{j\}}(\mathbf{g}) = -10(N_g - 1) \sum_{i \neq j} g_i^2 - 10 \cos(2\pi g_i)$, thereby removing one selected dimension; and $\alpha = \frac{\sqrt{N_g M}}{|\hat{\mathcal{M}}| - 1}$ is scaled such that diversity and quality have comparable importance. Taking the archive-summed fitness rather than archive-averaged fitness implies a bonus for larger archives. In a second condition, called **QD-Meta Translation**, the meta-fitness aims to provide high-quality solutions despite various translations to the genotype changing the location of the global optimum. For QD-Meta

Translation, the meta-fitness of an archive \mathcal{M} is the average of the archive-summed fitness across different translations:

$$\mathcal{F}_t(\mathcal{M}) = -\frac{10}{|\mathcal{T}|N_g} \left(\sum_{T \in \mathcal{T}} \sum_{g \in \mathcal{M}} \sum_{i=1}^{N_g} T(g_i)^2 - 10 \cos(2\pi T(g_i)) \right), \quad (5)$$

where \mathcal{T} is a set of randomly generated linear transformations of the type $T(x) \leftarrow ax + b$ where $b \in [-0.50, 0.50]$ represents a small to medium-sized shift of the fitness landscape and $a \in [-1.10, -0.91] \cup [0.91, 1.10]$ represents a 10% expansion or shrinkage of the landscape, with or without a sign change. To reduce variance across meta-evolution while sampling different values of a and b , we sample transformations based on 8 non-overlapping ranges determined by three boolean features: the sign of a , whether a shrinks or expands x , and the sign of b . So, for example, T_1 has $a \sim U(-1.10, -1)$, $b \sim U(-0.50, 0)$, T_2 has $a \sim U(-1.10, -1)$, $b \sim U(0, 0.50)$, T_3 has $a \sim U(1, 1.10)$, $b \sim U(-0.50, 0)$, etc. For each range there are two repetitions, yielding 16 random translations in each meta-fitness evaluation.

b) Test phase: The test phase assess rapid adaptation to fitness landscape changes in no more than 100 function evaluations of random search over the behavioural archive, randomly selecting solutions without replacement. Two distinct and independently run tests are set up which compared to the meta-fitness functions are conceptually similar but have a completely different distribution of optimal solutions.

In the **Dimension test**, one removes two selected dimensions of the genotype from the Rastrigin computation and replaces it with high-frequency sinusoidal noise with the same amplitude (i.e. 10) as the cosine in the original Rastrigin):

$$\tilde{f}_{\mathcal{T}}(\mathbf{g}) = -10(N_g - |\mathcal{J}|) \left(\sum_{i \notin \mathcal{J}} g_i^2 - 10 \cos(2\pi g_i) - \sum_{i \in \mathcal{J}} 10 \sin(6\pi g_i) \right), \quad (6)$$

where \mathcal{J} is a chosen index set. While conceptually similar to QD-Meta Dimension, the key difference is that two dimensions are altered and that the altered dimensions are replaced by a high-amplitude sine wave rather than completely being removed.

In the **Translation test**, one transforms the genotype using a linear transformation $g_i \leftarrow ag_i + b$ for all $i \in \{1, \dots, N_g\}$ before computing the 20-dimensional Rastrigin (see Eq. 3). In the test set, 120 unique index sets are generated, each of size $|\mathcal{J}| = 2$. Twelve settings of the slope, a , are generated by applying, or not applying, shrinkage and reversal to $a' = \{0.01, 0.25, 0.5\}$. The settings using $a' = 0.25$ and $a' = 0.50$, with or without shrinkage or reversal, are far out-of-distribution compared to the QD-Meta Translation's meta-fitness. The intercept, b , is chosen in $\pm\{0.01, 0.25, 0.50, 0.75, 1.0\}$. Therefore, only 8 out of 120 (a, b) -values are in-distribution for QD-Meta Translation and values close to these are only rarely chosen during meta-fitness evaluation.

C. Hexapod robot locomotion benchmark

To evaluate the QD systems on a realistic application, they are compared on the RHex hexapod robot platform (see Figure

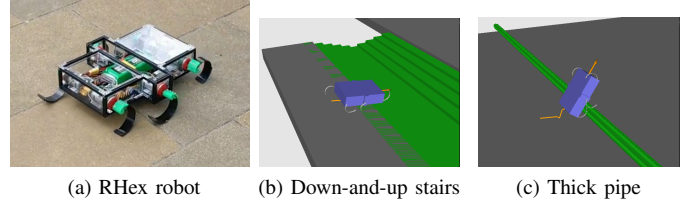


Fig. 2. The RHex hexapod robot platform: (a) the physical robot, on which the simulation is based; (b) the down-and-up stairs obstacle course; (c) the thick pipe obstacle course. For the full set of obstacle courses, see Fig. S2 of the Supplemental Materials.

2) [34], which is of interest because RHex can move across difficult terrains at high speeds. For QD-Meta, we independently run 4 different conditions based on two meta-objectives and two types of base-behavioural spaces per meta-objective. The meta-objectives are based on the performance after damaging one of the RHex robot's legs or after introducing an obstacle into the environment. The base-behavioural space is based on either feature-sets, which are traditionally used in quality-diversity meta-evolution, or the trajectory of the robot, as is traditionally used in AURORA and CVT-MAP-Elites. In the test phase, we devise two independent test scenarios, namely damages to the robot's legs as well as obstacle-courses. These test scenarios are not presented during the evolutionary phase and therefore represent adaptation to a priori unknown events.

a) Evolution phase: The RHex robot is simulated using the DART (Dynamic Animation and Robotics Toolkit) physics engine [35]. The task of the Rhex robot is to walk in a straight line and the fitness function f is the total distance the robot has moved forward within a time span of 5 s, on a flat plain surface where the robot faces no damages or obstacles. The control cycle of the locomotion controller is 5 ms.

Each leg of the robot is controlled by a Buehler clock, which alternates between a stance phase where the robot leg touches the ground and a swing phase where the leg rotates above ground [34], [36]. The genotype \mathbf{g} is comprised of 24 parameters, including: the period of the clocks, T , between 0.33-1 s (clock speed within 1-3 Hz; parameter 1); for each leg its duty-factor, the proportion of time that the leg touches the ground (parameters 2-7); the stance angle in $[0, \pi]$, the angle in which the robot leg touches the ground (parameters 8-13); the stance offset in $[-\pi/4, \pi/4]$, the angular offset to the leg's stance phase (parameters 14-19); the phase offset in $[0, T/2]$, the time lag with which the leg touches the ground when compared to leg 1 (parameters 20-24; reference leg 1 is excluded as parameter). The resulting 24-dimensional genotypic space is discretised with a step size of 0.025 and therefore whenever a gene is mutated, the mutation operator generates random increments or decrements of one step. The RHex robot experiments involve computationally expensive simulations,⁴ so the budget is limited to 12 million function evaluations and only 4,096 solutions are allowed.

Settings common to all algorithms are as follows. The base-behavioural space is a 50-dimensional space in $[0, 1]^{50}$ repre-

⁴Experiments are conducted on Intel Xeon Gold 6138 (2.00GHz) CPUs. While for Rastrigin, a single run requires only one CPU run for at most 20h, the RHex robot experiment runs require 40 CPUs for around 300h.

sending 10-step trajectories of the robot in terms of Cartesian xy -coordinates and the roll-pitch-yaw coordinates⁵. This choice serves two purposes: 1) these features are comparable to the typical AURORA setup thereby allowing a fair comparison; and 2) while previous work on QD-Meta used hand-crafted feature-sets, it is not clear whether such feature-sets are required or whether trajectories similar to those used in AURORA are sufficient. This would be important to know because the construction of feature-sets often requires additional prior knowledge (as mentioned in [20]), resulting in a potential trade-off between applicability and the customisation.

Settings specific to particular algorithms are as follows. For CVT-MAP-Elites, feasible trajectories must be sampled before forming centroids since a large part of the base-behavioural space may not provide a feasible solution; unlike the Rastrigin function optimisation, where all behaviours are feasible, the RHex robot may have a validly formed genotype without adhering to safety criteria. To obtain a suitable dataset for the feasible behaviour space, we sample genotypes until there are 100,000 valid robot trajectories that do not violate the safety measure (i.e. the RHex robot did not turn over). For QD-Meta, we additionally include the original feature-sets (see [22]), which combine: 1) the linear velocity in the 3 Cartesian axes (3D); 2) the frequencies of the body orientation being significantly negative or positive in the 3 orientation axes of roll, pitch, and yaw (6D); and 3) for each leg, the duty factor, i.e. the proportion of time it touches the ground. The different base-behaviour spaces will be distinguished by the suffix **Feature-sets** versus **Trajectory**. Two distinct QD-Meta conditions are included according to two different meta-objectives. In the **QD-Meta Damage** condition, the meta-objective is based on the average performance of the archive \mathcal{M} on a set of damages injected to the robot's legs (one-by-one):

$$\mathcal{F}_{\text{damage}}(\mathcal{M}) = \frac{1}{6|D||\mathcal{M}|} \sum_{\mathbf{g} \in \mathcal{M}} \sum_{d \in D} \sum_{l=1}^6 f(\mathbf{g}; d(l)), \quad (7)$$

where D defines a set of two damage-types, selected randomly from the full damage set at the start of the evolutionary experiment; and $f(\mathbf{g}; d(l))$ computes the fitness of the genotype \mathbf{g} when a damage of type d is applied to leg l of the RHex robot. The full damage set, from which the meta-fitness is constructed, comprises four types of damages to the RHex robot's legs: 1) leg-removal, which completely removes one leg; 2) leg-shortening, which shortens one leg significantly, usually preventing it from touching the floor; 3) blocked-joint, in which a joint cannot move; or 4) passive-joint, in which the affected joint cannot be controlled but can move passively. In the **QD-Meta Damage** condition, the meta-objective is based on the average performance of the archive \mathcal{M} on a set of obstacles:

$$\mathcal{F}_{\text{obstacle}}(\mathcal{M}) = \frac{1}{|O||\mathcal{M}|} \sum_{\mathbf{g} \in \mathcal{M}} \sum_{o \in O} f(\mathbf{g}; o), \quad (8)$$

⁵A tight normalisation in $[0, 1]^{50}$ is obtained by dividing at each step the relative change from the previous step by the maximal deviations; for example, to compute x_2 as a feature, $x_2 \leftarrow 0.5 + 0.5 \frac{x_2 - x_1}{\Delta x}$ is computed where Δx is an empirical estimate of the maximal absolute value of x .

where O defines a set of 5 obstacles selected at random out of the full obstacle set at the start of the evolutionary experiment. The full obstacle set comprises 9 obstacle courses (see Figure S2 in Supplemental Materials): a) a large sphere interrupting the normal trajectory; b) a continuing upward stairs; c) a slope; d) a large pile of rubble forming a rough terrain; e) downward stairs followed by upward stairs; f) several small pipes evenly spaced across the forward trajectory; g) a ditch; h) a thick pipe; and i) a thin pipe. To avoid selecting problems that are not solvable, the obstacle courses were adjusted to meet a target optimal performance between 3 m and 5 m, which was confirmed by obstacle-specific runs with a traditional EA.

b) *Test phase*: In the test phase, the archives are assessed on obstacles and damages not known a priori by any of the algorithms. The **Obstacle test** assesses rapid performance recovery on obstacles not experienced during meta-evolution while the **Damage test** assesses rapid performance recovery on damages not experienced during meta-evolution. In both cases, recovery is done by performing random search (without replacement) in no more than 50 function evaluations. To assess the overall robustness of solutions, we also perform a **Generalisation test** in which *all* controllers are assessed for their average performance over *all* environments.

IV. RESULTS

This section analyses the comparative performance of QD-Meta and the baseline algorithms on the function optimisation and RHex robot experiments. For each benchmark, we evaluate the evolutionary development, including the number of solutions in the archive and their fitness over time, as well as their performance on the consequent test. A further qualitative analysis probes into the behavioural differences observed when comparing different meta-objectives.

A. Rastrigin function optimisation

In the evolution phase, the quality-diversity statistics can be observed from Figure 3. All QD algorithms yield a high number of solutions although there is a difference between CVT-MAP-Elites, which yields just below 3,000 solutions, and AURORA and QD-Meta conditions, which yield about 10,000 solutions. The difference is explained by AURORA setting 10,000 as an explicit target for the proportional container size control and by QD-Meta conditions having a meta-fitness bonus for the number of solutions. All algorithms converge to a global fitness of around -70 . The largest difference observed is that QD-Meta scores highest on the average fitness, with a score of -100 , followed by CVT-MAP-Elites with a score of -170 , and finally AURORA, -300 . The higher average fitness of QD-Meta conditions is explained by the meta-fitness bonus for covering significant peaks of the fitness landscape (i.e. those peaks that can become top-performers after translation or dimensionality change) rather than covering the behaviour space evenly. The population-based methodology of evolving multiple feature-maps as meta-genotypes is validated by an analysis comparing the results for QD-Meta Translation with different population sizes $\lambda \in \{1, 2, 5, 10, 20\}$ (see Fig. S1 in Supplemental Materials; dynamic parameter control disabled

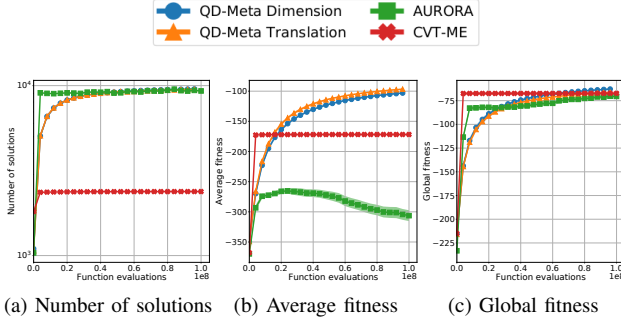


Fig. 3. Quality-diversity statistics (Mean \pm SE) of the different included QD algorithms across 20 replicates on the Rastrigin function optimisation, including (a) the total number of solutions in the archive; (b) the average fitness across the archive; and (c) the maximal fitness across the archive. For QD-Meta, Mean and SE statistics are aggregated across replicates and the different archives within the meta-population.

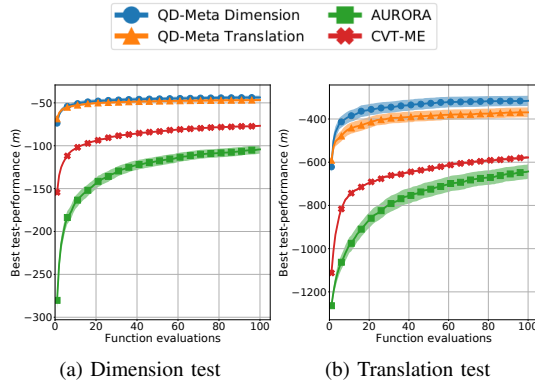


Fig. 4. Test performance (Mean \pm SE) of the different included QD algorithms across 20 replicates of the Rastrigin function optimisation benchmark. The y -axis shows the best solution so far after performing random search over the behavioural archive for the number of function evaluations indicated on the x -axis. For each replicate of QD-Meta, the archive with the highest meta-fitness at the end of meta-evolution is chosen.

for simplicity), which shows that a population size of 1 has low performance on meta-fitness and QD metrics.

In the test phase, a random search across the QD archives allowing at most 100 function evaluations (see Figure 4) demonstrates that QD-Meta algorithms significantly outperform other algorithms. Both meta-fitness conditions, QD-Meta Dimension and QD-Meta Translation, have a beneficial effect on generalisation, allowing rapid adaptation on both the Dimension test and the Translation test.

B. Hexapod robot locomotion

In robotics applications, a key challenge is to adapt to environmental changes within a limited number of trials without trying out low-quality controllers that may cause performance loss or even safety violations. The results of the RHex robot study highlight how QD-Meta can provide archives with a lower number of solutions but a much higher average performance in unforeseen circumstances.

As observable in Figure 5, rather than increasing the archive coverage, QD-Meta actually decreases the archive coverage over time, yielding around 100 controllers (compare to 3,000 to 4,000 controllers for AURORA and CVT-MAP-Elites). At the same time, QD-Meta increases the average fitness

in the normal environment up to between 6 and 8 meters, outperforming in this regard CVT-MAP-Elites (5m) and AURORA (2m). This points out the ability of QD-Meta to determine the optimal archive size for the specified meta-objective: since the behaviour space is optimised to allow high average performance of controllers across different obstacles and damages, QD-Meta results in smaller archives with only high-quality solutions. Given that the meta-fitness is biased for average fitness across a variety of environments related to the normal evolutionary environment, the higher average fitness in the normal environment is perhaps unsurprising. However, further analysis shows that all algorithms have comparable global fitness, though a small advantage of QD-Meta and AURORA (both around 9.5m) over CVT-MAP-Elites (9.4m) can be observed; therefore, QD-Meta's lower number of solutions and bias for average performance does not come at the cost of global fitness in the normal evolutionary environment. Furthermore, QD-Meta's smaller archives with high-quality solutions are beneficial for adapting to a priori unknown environments in two ways. First, QD-Meta archives have improved performance on the Generalisation test compared to CVT-MAP-Elites and AURORA (see Table II). This improvement is seen whether or not the damages or obstacles were experienced in the meta-fitness evaluations. Second, in both the Obstacle and Damage test, when performing a search across the archive for the best controller for a particular environment, the QD-Meta archives require only a few function evaluations to reach a near-optimal solution (see Figure 6). However, a trade-off of the smaller archives is that QD-Meta archives do not always result in highest global test performance; with a large number of solutions, even a low-quality archive can eventually yield a high maximum as long as there is one solution well suited to that particular environmental change.

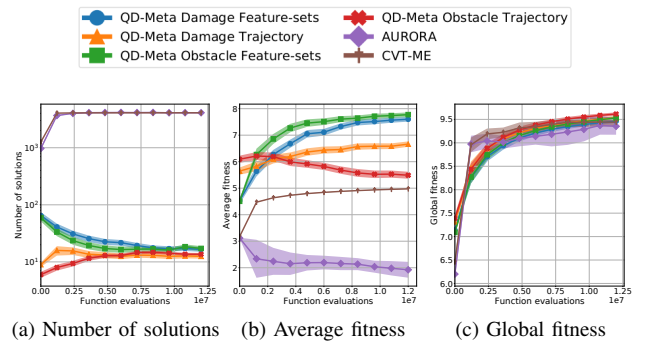


Fig. 5. Quality-diversity statistics (Mean \pm SE) of the different included QD algorithms across 4 replicates on the RHex robot platform, including (a) the total number of solutions in the archive; (b) the average fitness across the archive; and (c) the maximal fitness across the archive. For QD-Meta, Mean and SE statistics are aggregated across replicates and the different archives within the meta-population.

A qualitative analysis using slow-motion video material (see <http://tiny.cc/QD-Meta>) demonstrates the different behaviour of QD-Meta Obstacle conditions compared to QD-Meta Damage conditions. In particular, the behaviours evolved by QD-Meta Damage appear to less balanced with more roll motion. In the

exemplary videos included for QD-Meta Damage (see videos with QD-Meta-Damage prefix), one of the legs has its stance phase facing upwards, which is comparable to disabling the leg as if it is damaged. In the exemplary videos included for QD-Meta Obstacle (see videos with QD-Meta-Obstacle prefix), one can observe jumpy behaviours, which help to navigate over obstacles such as pipes and stairs. The mentioned QD-Meta Damage behaviours are rather specific in comparison; as they seem to disable one or more legs, they result in shaky behaviours that particularly perform well when the disabled leg is damaged but that do not generalise well to other environments. This explains why QD-Meta Damage does not generalise as well as QD-Meta Obstacle to new obstacles or new damages (see Table II). In particular, one can observe the larger generalisation drop from Training to Test for Damage conditions and the consistent top ranking of QD-Meta Obstacle Feature-sets.

A further analysis reveals the nature of the benefits of dynamic parameter control with RL. A meta-fitness comparison (see Fig. S4 in Supplementary materials) finds that the benefit is typically limited although the Obstacle Trajectory condition with RL obtains larger and smoother meta-fitness improvements than the version without RL. By further analysing each run separately and comparing the evolution of the meta-fitness to the controlled parameter (i.e. the generations per meta-generation; see Fig. S5 in Supplementary), a pattern is observed in which the number of generations is high initially, declines sharply, and finally levels off or increases slightly as evolution progresses. This is in line with the hypothesis that initially the algorithm should generate a large number of behaviourally diverse and high-performing solutions – using ME iterations – to be able to get reliable evaluations of the behaviour space – using meta-fitness evaluations. Low (resp. high) meta-fitness after the initial phase results in a low (resp. high) ratio of generations per meta-generation. This indicates that: for regions around the global optimum, high-reliability evaluations are required; for local optima, evaluating many feature-maps helps to rapidly escape to a higher-performing region.

V. DISCUSSION

This paper demonstrates QD-Meta as a promising framework for the automated design of low-dimensional behaviour spaces for QD optimisation. We present results showing that QD-Meta archives improve rapid adaptation to fitness landscapes not experienced during evolution, further establishing QD-Meta as a widely applicable QD framework with state-of-the-art performance. QD-Meta compares favourably to state-of-the-art QD algorithms for automated behaviour spaces, namely AURORA [20] and CVT-MAP-Elites [18], not only in adaptation to new environments but also on traditional QD metrics; that is, the average fitness and the number of solutions are improved when they are rewarded for in the meta-fitness. Additionally, the archives evolved by QD-Meta further confer unique qualitative benefits depending on the meta-objective: behaviours adapted to various damages are quite distinct and specialised compared to behaviours adapted to various obstacles. Furthermore, in experiments on the Rastrigin function, the

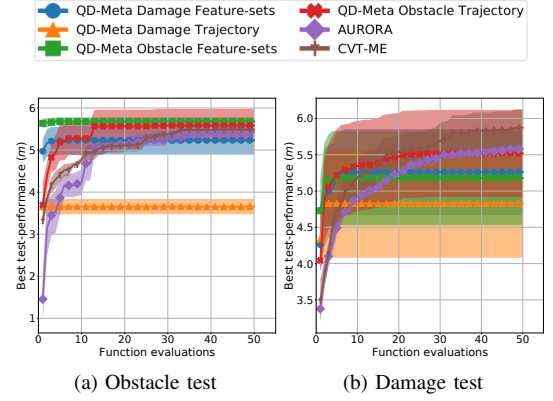


Fig. 6. Test performance (Mean \pm SE) of the different included QD algorithms on the RHex robot platform across 4 replicates. The y -axis shows the best solution so far after performing random search without replacement over the behavioural archive for the number of function evaluations indicated on the x -axis. For comparability, the lines are continued when the archive is exhausted, even though no further function evaluations are performed. For each replicate of QD-Meta, the archive with the highest meta-fitness at the end of meta-evolution is chosen.

population-based nature of QD-Meta is shown to be essential for meta-fitness as well as standard QD metrics. With the exception of the global fitness, which requires a relatively large meta-population of 10 to 20 meta-individuals, a meta-population size of 2 is sufficient to achieve this effect. Finally, it is shown that hand-crafted feature-sets can usually be replaced by high-dimensional trajectories without performance loss although these require more evolutionary function evaluations due to the high dimensionality of the meta-genotype.

The findings on function optimisation show that QD-Meta allows to evolve archives that are robust to dimensionality decreases and translations. This is of interest on a wide variety of applications as the data that an algorithm is trained on does not always reflect the real distribution of the data or the function may even change dynamically over time. While both meta-fitness formulations in the Rastrigin have a positive effect on both tests, the effect of the drop of dimensionality of the inputs is stronger. One interpretation is that performing dimensionality reductions in the meta-fitness has an effect on the feature-map similar to Dropout [37], which improves generalisation in neural networks by randomly dropping out the activation of some of its neurons, typically including input neurons.

Traditionally, quality-diversity optimisation has assumed that a large number of solutions is better. However, smaller archives can provide key benefits for adaptation and greater transparency. In this context, QD-Meta has the unique benefit of being able to tailor the behaviour space, including the number of solutions as well as the behavioural features, to a meta-objective rather than being unsupervised [20] or unadaptive [18]. This benefit can be exploited to suit the domain. If an end-user is interested in multi-modal function optimisation domains, and in particular covering the peaks of the fitness function despite dynamic changes, then the meta-fitness could provide a bonus for archive size as in the present Rastrigin experiments. In domains such as multi-legged robot locomotion, the same gait can overcome different types of obstacles near-optimally, and

TABLE II

GENERALISATION TEST PERFORMANCE (MEAN \pm SD) ON THE TRAINING AND TEST ENVIRONMENTS OF THE RHEX ROBOT PLATFORM. BOLD INDICATES THE BEST PERFORMANCE FOR THE GIVEN ENVIRONMENT. PERFORMANCE IS THE AVERAGE ACROSS THE SOLUTIONS IN THE ARCHIVE AND THE DIFFERENT INCLUDED TEST ENVIRONMENTS. MEAN AND STANDARD DEVIATION ARE AGGREGATED ACROSS 4 REPLICATES. FOR EACH REPLICATE OF QD-META, THE ARCHIVE WITH THE HIGHEST META-FITNESS AT THE END OF META-EVOLUTION IS CHOSEN. UNDERLINE INDICATES THE OBSTACLES OR DAMAGES ARE EXPERIENCED DURING META-EVOLUTION.

Environment	Condition				Baseline	
	QD-Meta				AURORA	CVT-ME
	Damage Feature-sets	Damage Trajectory	Obstacle Feature-sets	Obstacle Trajectory		
Training-Damages	6.779 \pm 0.89	6.882 \pm 1.06	6.876 \pm 1.14	5.093 \pm 1.26	3.072 \pm 0.19	4.451 \pm 0.25
Training-Obstacles	3.789 \pm 0.77	3.017 \pm 0.90	4.059 \pm 0.66	3.895 \pm 0.45	2.060 \pm 0.11	2.660 \pm 0.11
Test-Damages	4.491 \pm 0.99	4.193 \pm 0.66	4.782 \pm 0.88	4.063 \pm 0.64	3.404 \pm 0.11	3.422 \pm 0.24
Test-Obstacles	3.628 \pm 0.50	2.733 \pm 0.61	4.161 \pm 0.72	2.831 \pm 0.20	2.400 \pm 0.26	2.642 \pm 0.16

therefore only a few solutions may be required so a meta-fitness bonus for archive size is not necessary. Indeed, in the RHex robot study, QD-Meta demonstrates the benefits of having a lower number of solutions with strong generalisation across different environments: 1) all of the solutions have a high performance even when the environment changes so there is often no need for adaptation; 2) all the controllers can safely be deployed during adaptation without too much risk of low performance or otherwise unsafe behaviours; and 3) as there is only a need to try a dozen solutions, the archives allow rapid adaptation. A further benefit of smaller archives may be in transparency and verification: if the number of solutions in the archive is limited, then the end-user can easily verify and assess the solutions to prevent undesirable consequences from their application. That said, if the number of trialled solutions is not of critical concern but instead the maximal performance, then a meta-fitness bonus for larger archives would be preferable.

In previous work QD-Meta relied on hand-crafted feature-sets. While in most robotics applications feature-sets can be given to the user or constructed with only minor effort, the trajectory of the robot is a generic choice that requires limited to no prior knowledge; for this reason it has been the base-behavioural space of choice in methods such as AURORA and CVT-MAP-Elites. The present study shows that while there is a benefit to feature-sets, QD-Meta can achieve a comparable performance using the robot’s trajectories. A difficulty observed is that the meta-optimisation problem becomes more challenging due to the increased dimensionality and the authors recommend research into scaling up the approach to even higher-dimensional base-behavioural spaces, such as (sequences of) raw pixel data. To improve database storage capacity as well as the optimisation of the feature-map in the meta-genotypic space, one suggested approach is to first apply traditional dimensionality reduction methods [38] to yield a low-dimensional “latent” base-behavioural space which is then provided as input to the feature-map.

Evolving a meta-population of QD archives rather than a single meta-individual is shown to be essential for the search for high meta-fitness feature-maps, with meta-fitness and QD metrics improving as the meta-population size is increased to 10. Two explanations are proposed for its benefits. First, population-based optimisation algorithms are less prone to get stuck in local optima due to being distributed across the search space; this is crucial in large multi-modal search spaces such

as the 522-dimensional space of feature-maps in QD-Meta. Second, each meta-individual comes with its own behaviour space, affecting which elites are selected for reproduction, and therefore QD-Meta generates new solutions from different sources. Beyond the earlier meta-evolution works [22], [23], a similar reasoning can be found in recent works in the literature. Multi-emitter MAP-Elites uses different “emitters”, each of which provide a different reproduction operator based on different heuristics to explore the search space, for example based on fitness or a random direction in the behaviour space [39]. Multi-container AURORA maintains a number of feature-maps at the same time to yield improvements to the QD-score, which is the $[0, 1]$ -normalised archive-summed fitness [40]. These improvements were minor compared to the average fitness improvements in our Rastrigin function optimisation results. Future research is advised to compare and integrate these strategies to yield new insights into QD optimisation.

VI. CONCLUSION

This paper demonstrates Quality-Diversity Meta-evolution as a promising framework for quality-diversity optimisation. Quality-diversity metrics as well as tests on changed fitness landscapes demonstrate that QD-Meta outperforms CVT-MAP-Elites as well as AURORA, two state-of-the-art algorithms for automated behaviour spaces from high-dimensional trajectories. The study shows improved adaptation to linear transformations and dimensionality changes in function optimisation as well as adaptation of multilegged locomotion gaits to various robot damages and various obstacle courses. Compared to existing methods, QD-Meta derives its adaptation capability to customising the behaviour space to a generalisation-based meta-objective. The study also demonstrates that QD-Meta can yield similar benefits with high-dimensional observation trajectories without requiring the user to hand-craft feature-sets. For future research, an investigation into further scalability to even higher-dimensional base-behaviour spaces (e.g. raw pixel data streams) is advised, as is a study integrating and comparing multi-emitter and multi-container QD algorithms to QD-Meta variants.

ACKNOWLEDGEMENTS

This work has been supported by the Engineering and Physical Sciences Research Council under the New Investigator Award grant, EP/R030073/1, and the UKRI Trustworthy

Autonomous Systems Hub, EP/V00784X/1. The authors acknowledge the use of the IRIDIS High Performance Computing Facility.

REFERENCES

- [1] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Combining convergence and diversity in evolutionary multiobjective optimization," *Evolutionary Computation*, vol. 10, no. 3, pp. 263–282, 2002.
- [2] D. Gupta and S. Ghafir, "An Overview of methods maintaining Diversity in Genetic Algorithms," *International Journal of Emerging Technology and Advanced Engineering*, vol. 2, no. 5, pp. 56–60, 2012.
- [3] R. K. Ursem, "Diversity-guided evolutionary algorithms," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 2439, pp. 462–471, 2002.
- [4] B. M. Ginley, J. Maher, C. O. Riordan, and F. Morgan, "Adaptive Crossover, Mutation, and Selection," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 5, pp. 692–714, 2011.
- [5] S. W. Mahfoud, "Nicheing methods for genetic algorithms," Ph.D. dissertation, 1995.
- [6] A. Toffolo and E. Benini, "Genetic diversity as an objective in multi-objective evolutionary algorithms," *Evolutionary Computation*, vol. 11, no. 2, pp. 151–167, 2003.
- [7] J.-B. Mouret and S. Doncieux, "Using behavioral exploration objectives to solve deceptive problems in neuro-evolution," *GECCO 2009 – Proceedings of the 2009 Genetic and Evolutionary Computation Conference*, pp. 627–634, 2009.
- [8] F. J. Gomez, "Sustaining diversity using behavioral information distance," in *GECCO 2009: Proceedings of the 2009 Genetic and Evolutionary Computation Conference*, 2009, pp. 113–120.
- [9] J.-B. Mouret and S. Doncieux, "Overcoming the bootstrap problem in evolutionary robotics using behavioral diversity," in *CEC 2009: Proceedings of the 2009 IEEE Congress on Evolutionary Computation*, 2009, pp. 1161–1168.
- [10] J.-B. Mouret, "Behavioral diversity measures for Evolutionary Robotics," in *CEC 2010 – Proceedings of the 2010 IEEE Congress on Evolutionary Computation*, 2010, pp. 1303–1310.
- [11] J. Lehman and K. O. Stanley, "Abandoning objectives: Evolution through the search for novelty alone," *Evolutionary Computation*, vol. 19, no. 2, pp. 189–222, 2011.
- [12] J.-B. Mouret and J. Clune, "Illuminating search spaces by mapping elites," *arXiv preprint arXiv:1504.04909v1*, pp. 1–15, 2015.
- [13] J. Lehman and K. O. Stanley, "Evolving a Diversity of Creatures through Novelty Search and Local Competition," in *GECCO 2011 – Proceedings of the 2011 Genetic and Evolutionary Computation Conference*, 2011, p. 1408.
- [14] J. Nordmoen, K. O. Ellefsen, and K. Glette, "Combining MAP-Elites and Incremental Evolution to Generate Gaits for a Mammalian Quadruped Robot," in *EvoApplications 2018, Lecture Notes on Computer Science 10784*, K. Sim and P. Kaufmann, Eds., 2018, pp. 719–733.
- [15] A. Cully, J. Clune, D. Tarapore, and J.-B. Mouret, "Robots that can adapt like animals," *Nature*, vol. 521, no. 7553, pp. 503–507, may 2015.
- [16] D. M. Bossens and D. Tarapore, "Rapidly adapting robot swarms with Swarm Map-based Bayesian Optimisation," in *ICRA 2021 – Proceedings of the 2021 International Conference on Robotics and Automation*, 2021.
- [17] E. Meyerson, J. Lehman, and R. Mäkiulainen, "Learning Behavior Characterizations for Novelty Search," in *GECCO 2016: Proceedings of the 2016 Genetic and Evolutionary Computation Conference*, 2016, pp. 149–156.
- [18] V. Vassiliades, K. Chatzilygeroudis, and J. B. Mouret, "Using Centroidal Voronoi Tessellations to Scale Up the Multidimensional Archive of Phenotypic Elites Algorithm," *IEEE Transactions on Evolutionary Computation*, vol. 22, no. 4, pp. 623–630, 2018.
- [19] A. Cully, "Autonomous skill discovery with quality-diversity and unsupervised descriptors," in *GECCO 2019 – Proceedings of the 2019 Genetic and Evolutionary Computation Conference*, 2019, pp. 81–89.
- [20] L. Grillotti and A. Cully, "Unsupervised Behaviour Discovery with Quality-Diversity Optimisation," *IEEE Transactions on Evolutionary Computation*, vol. 25, no. 4, pp. 1–13, 2021.
- [21] G. Paolo, A. Laflaquiere, A. Coninx, and S. Doncieux, "Unsupervised Learning and Exploration of Reachable Outcome Space," *ICRA 2020 – Proceedings of the 2020 International Conference on Robotics and Automation*, pp. 2379–2385, 2020.
- [22] D. M. Bossens, J.-B. Mouret, and D. Tarapore, "Learning behaviour-performance maps with meta-evolution," in *GECCO 2020 – Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, 2020, pp. 49–57.
- [23] D. M. Bossens and D. Tarapore, "On the use of feature-maps and parameter control for improved quality-diversity meta-evolution," *arXiv preprint arXiv:2105.10317*, 2021.
- [24] N. Hansen, "The CMA Evolution Strategy: A Comparing Review," *Towards a New Evolutionary Computation*, vol. 102, no. 2006, pp. 75–102, 2007.
- [25] —, "The CMA Evolution Strategy: A Tutorial," *arXiv preprint arXiv:1604.00772*, pp. 1–39, 2016.
- [26] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function," *Neural Networks*, vol. 6, no. 6, pp. 861–867, 1993.
- [27] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.
- [28] G. Karafotias, A. E. Eiben, and M. Hoogendoorn, "Generic parameter control with reinforcement learning," *GECCO 2014 – Proceedings of the 2014 Genetic and Evolutionary Computation Conference*, pp. 1319–1326, 2014.
- [29] J. Nordmoen, E. Samuelsen, K. O. Ellefsen, and K. Glette, "Dynamic mutation in MAP-Elites for robotic repertoire generation," in *2018 Conference on Artificial Life: A Hybrid of the European Conference on Artificial Life (ECAL) and the International Conference on the Synthesis and Simulation of Living Systems (ALIFE)*, 2018, pp. 598–605.
- [30] G. Karafotias, M. Hoogendoorn, and A. E. Eiben, "Parameter Control in Evolutionary Algorithms: Trends and Challenges," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 2, pp. 167–187, 2015.
- [31] W. T. Uther and M. M. Veloso, "Tree based discretization for continuous state space reinforcement learning," *AAAI 94: Proceedings of the 1994 National Conference on Artificial Intelligence*, no. 1994, pp. 769–774, 1998.
- [32] G. A. Rummery and M. Niranjan, "Online Q-learning Using Connectionist Systems," Cambridge University Engineering Department, Tech. Rep., 1994.
- [33] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction*. MIT Press, 2018.
- [34] U. Saranli, M. Buehler, and D. E. Koditschek, "RHex: A Simple and Highly Mobile Robot," *International Journal of Robotics Research*, vol. 20, no. 7, pp. 616–631, 2001.
- [35] J. Lee, M. X. Grey, S. Ha, T. Kunz, S. Jain, Y. Ye, S. S. Srinivasa, M. Stilman, and C. Karen Liu, "DART: Dynamic Animation and Robotics Toolkit," *The Journal of Open Source Software*, vol. 3, no. 22, p. 500, 2018.
- [36] J. Seipel and P. Holmes, "A simple model for clock-actuated legged locomotion," *Regular and Chaotic Dynamics*, vol. 12, no. 5, pp. 502–520, 2007.
- [37] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [38] C. O. S. Sorzano, J. Vargas, and A. P. Montano, "A survey of dimensionality reduction techniques," *arXiv preprint arXiv:1403.2877*, pp. 1–35, 2014.
- [39] A. Cully, "Multi-emitter MAP-elites: Improving quality, diversity and data efficiency with heterogeneous sets of emitters," in *GECCO 2021 – Proceedings of the 2021 Genetic and Evolutionary Computation Conference*, 2021, pp. 84–92.
- [40] L. Cazenille, "Ensemble Feature Extraction for Multi-Container Quality-Diversity Algorithms," in *GECCO 2021 – Proceedings of the 2021 Genetic and Evolutionary Computation Conference*, 2021, pp. 75–83.